Multimodality imaging-based response prediction and monitoring to improve clinical management of gastrointestinal stromal tumours

Ylva A. Weeda

Master thesis
Technical Medicine | Imaging & Intervention

Thesis committee
Prof. dr. Jos A. van der Hage
Prof. dr. Lioe-Fee de Geus-Oei
Dr. Willem Grootjans
Dr. Jifke F. Veenland
Drs. Gijsbert M. Kalisvaart

Multimodality imaging-based response prediction and monitoring to improve clinical management of gastrointestinal stromal tumours

Ylva Anouk Weeda

Student number : 4556038

3rd of March 2023

Thesis in partial fulfilment of the requirements for the joint degree of Master of Science in

Technical Medicine - Imaging & Intervention

Leiden University | Delft University of Technology | Erasmus University Rotterdam

Master thesis project (TM30004 | 35 ECTS)

Dept. of Radiology, Section of Nuclear medicine,
Leiden University Medical Center

May 2022 – March 2023

Supervisor(s):

Prof. Dr. Jos A. van der Hage

Prof. Dr. Lioe-Fee de Geus-Oei

Dr. Willem Grootjans

Drs. Gijsbert M. Kalisvaart

Thesis committee members:

Prof. Dr. Jos A. van der Hage (chair)

Prof. Dr. Lioe-Fee de Geus-Oei

Dr. Willem Grootjans,

Dr. Jifke F. Veenland

Drs. Gijsbert M. Kalisvaart

Preface and acknowledgments

With this thesis, I am closing off my time as a Technical Medicine student. After visiting the annual open days of the TU Delft and seeing the presentation on Technical Medicine, I knew this was the degree I wanted to pursue. Up to this day, I am still very happy about this decision!

When choosing a subject for my master thesis, the multidisciplinary treatment strategy of gastrointestinal stromal tumours piqued my interest. As a Technical Medicine student, I aim to bring departments together to combine medicine with engineering in the most patient-friendly way. This project fitted my wishes perfectly and included collaborations between the departments of radiology, pathology, surgical and medical oncology.

During this project, I have developed a growing interest for research. This interest was mainly fuelled by Willem. You have many innovative ideas, and I could not have asked for a better supervisor. I am looking forward to our future research and coffee meetings. I would also like to thank Timo. You have not only provided me with valuable input and great advice, but it also feels like I made a friend. It is truly inspiring to have a supervisor like Lioe-Fee. You are striving for more technical physicians in the field, and I felt honoured to have experienced your academic appointment as Medical Delta professor. Jos, your surgical experience is admiring, and I am very happy to have gotten a little peak into your field of work. All in all, you were so willing to exchange your knowledge, expertise and enthusiasm. You have created a very pleasant and stimulating learning and working environment, for which I am grateful! Jifke, your master courses have made me very excited about image acquisition, image processing and radiomics. It is an honour to have you join my thesis committee!

My colleagues at the nuclear medicine department have welcomed me with open arms and have made this experience so much more fun. It was extra special to have shared this experience with my fellow student and dear friend Gonnie.

I would also like to thank my friends, roommates and family for supporting me. I would especially like to show appreciation for my parents, as they have always been my biggest supporters and my source of inspiration. Last but not least, I want to praise Mats for making his way through all the medical jargon to aid me towards a more well-defined thesis project. I am incredibly grateful to have met you and I cannot wait for the exciting adventures that lie ahead of us.

Ylva Weeda

Summary

Gastrointestinal stromal tumours (GISTs) are rare mesenchymal neoplasms with a worldwide incidence of one or two per 100,000. The tumours affect the entire gastrointestinal tract, but most commonly the stomach and small intestine. Neoadjuvant tyrosine kinase inhibitors (TKI) are administered in a selection of GIST patients to attain size reduction of the primary tumour and improve chances of complete resection. Response monitoring in GISTs is complex due to the presence of intra- and intertumoral heterogeneity and lack of pathological criteria to define response. Nonetheless, it is of importance to evaluate the efficacy of TKI treatment at an early stage in order to optimise treatment. In particular, early cessation of ineffective treatment is of importance in these patients, preventing unnecessary side-effects and healthcare costs. Medical imaging plays an important role to non-invasively predict and monitor treatment response of GIST patients undergoing TKI-treatment. This thesis aims to improve understanding of contrast-enhanced computed tomography (CE-CT) and 2-deoxy-2-[^{18}F]fluoro-D-glucose ([^{18}F]FDG) PET/CT imaging parameters to allow prediction and monitoring of neoadjuvant treatment response in GIST patients.

In Chapter 1, the added value of CE-CT and [^{18}F]FDG-PET imaging for early prediction and monitoring of treatment response in GISTs is investigated by means of a systematic literature review. Results of this study show that heterogeneous enhancement patterns on baseline CE-CT imaging were considered to be predictive for high-risk GISTs, reflecting neovascularisation and the presence of necrosis. Current CE-CT radiographic response criteria (i.e., RECIST 1.1 and Choi) are still lacking sensitivity and are prone to errors when predicting or monitoring treatment response. Metabolic changes on [^{18}F]FDG-PET imaging seem to precede morphological changes in size in GIST lesions and were more strongly correlated with tumour response. Although CE-CT and [^{18}F]FDG-PET can aid in the prediction and monitoring in GIST patients, further research on cost-effectiveness is recommended.

Chapter 2 evaluates the efficacy of current radiological response criteria (RECIST 1.1, Choi and tumour volumetry) in predicting response to neoadjuvant systemic therapy, by comparing radiological response criteria with the achieved surgical benefit. Results show that size-based criteria (RECIST 1.1 and volumetry) accurately reflect surgical benefit in GIST patients treated with neoadjuvant systemic therapy (accuracy of 76.3% and 86.6% respectively) and are less prone to scanner and imaging protocol variabilities, when compared to the Choi criteria (68.4%).

In addition to volumetry, quantitative radiomic models using CE-CT and [^{18}F]FDG-PET imaging features were trained to predict response at baseline. Preliminary results of this study are described in Chapter 3. The radiomic models presented in this study generally had a poor performance and can therefore not yet be applied in a clinical setting. To improve performance and generalisability, future research should focus on a bigger patient population and harmonisation of acquisition protocols. The conclusion of this thesis supports the utilisation of tumour diameters for radiological response assessment, where RECIST 1.1 response criteria had an accuracy of 80.0% to correctly predict volumetric response after the first response follow-up CE-CT scan. When properly executed, these manual measurements could aid in early surgical decision making.

Table of contents

1

Early prediction and monitoring of treatment response in gastrointestinal stromal tumours by means of imaging: a systematic review

Early prediction and monitoring of treatment response in gastrointestinal stromal tumours by means of imaging: a systematic review

Ylva A. Weeda[1,*], Gijsbert M. Kalisvaart[1], Floris H.P van Velden[1], Hans Gelderblom[2], Aart J. van der Molen[3], Judith V.M.G. Bovee[4], Jos A. van der Hage[5], Willem Grootjans[1], Lioe-Fee de Geus-Oei[1,6,7]

[1]Department of Radiology, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands;
[2]Department of Medical Oncology, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands;
[3] Department of Radiology, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands;
[4]Department of Pathology, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands;
[5]Department of Surgical Oncology, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands;
[6]Biomedical Photonic Imaging Group, University of Twente, 7522 NB Enschede, The Netherlands
[7]Department of Radiation Science & Technology, Technical University of Delft, 2629 JB Delft, The Netherlands
* Correspondence: y.a.weeda@lumc.nl

Diagnostics 2022, 12, 2722. https://doi.org/10.3390/diagnostics12112722

Abstract: Gastrointestinal stromal tumours (GISTs) are rare mesenchymal neoplasms. Tyrosine kinase inhibitor (TKI) therapy is currently part of routine clinical practice for unresectable and metastatic disease. It is important to assess the efficacy of TKI treatment at an early stage to optimize therapy strategies and eliminate futile ineffective treatment, side effects and unnecessary costs. This systematic review provides an overview of the imaging features obtained from contrast-enhanced (CE)-CT and 2-deoxy-2-[^{18}F]fluoro-D-glucose ([^{18}F]FDG) PET/CT to predict and monitor TKI treatment response in GIST patients. PubMed, Web of Science, the Cochrane Library and Embase were systematically screened. Articles were considered eligible if quantitative outcome measures (area under the curve (AUC), correlations, sensitivity, specificity, accuracy) were used to evaluate the efficacy of imaging features for predicting and monitoring treatment response to various TKI treatments. The methodological quality of all articles was assessed using the Quality Assessment of Diagnostic Accuracy Studies, v2 (QUADAS-2) tool and modified versions of the Radiomics Quality Score (RQS). A total of 90 articles were included, of which 66 articles used baseline [^{18}F]FDG-PET and CE-CT imaging features for response prediction. Generally, the presence of heterogeneous enhancement on baseline CE-CT imaging was considered predictive for high-risk GISTs, related to underlying neovascularisation and necrosis of the tumour. The remaining articles discussed therapy monitoring. Clinically established imaging features, including changes in tumour size and density, were considered unfavourable monitoring criteria, leading to under- and overestimation of response. Furthermore, changes in glucose metabolism, as reflected by [^{18}F]FDG-PET imaging features, preceded changes in tumour size and were more strongly correlated with tumour response. Although CE-CT and [^{18}F]FDG-PET can aid in the prediction and monitoring in GIST patients, further research on cost-effectiveness is recommended.

Keywords: gastrointestinal stromal tumour; prediction; response monitoring; FDG-PET; radiomics; Tomography, X-ray Computed; personalized medicine

1. Introduction

Gastrointestinal stromal tumours (GISTs) are rare mesenchymal neoplasms affecting the entire gastrointestinal tract and are presumed to originate from the interstitial cells of Cajal [1,2]. About 80–90% of GISTs harbour kinase-activating mutations in either receptor tyrosine kinase protein (KIT) or platelet-derived growth factor receptor α (PDGRF-α) [3,4]. Complete surgical excision remains the only curative treatment option for GIST patients. Since GISTs are generally insensitive to radio- and chemotherapy, non-surgical treatment is limited to tyrosine kinase inhibitor (TKI) therapy. This targeted molecular therapy is part of routine clinical practice for unresectable and metastatic disease [5,6].

Adjuvant TKI treatment is used in high-risk GISTs to improve survival [7]. Unfortunately, due to the varying aggressive nature of GISTs, about one-third of the patients will relapse within three years after surgery with curative-intent [8]. For localized disease, TKI treatment can be given to attain size reduction of the primary tumour and improve chances of complete resection while maintaining an acceptable risk of complications [9,10]. About 20–25% of patients do not benefit from the neoadjuvant TKI treatment, as no complete or partial response is observed [11,12]. The rarity and complex biological nature of this disease, makes it difficult to differentiate between good and poor responders. For example, GISTs harbouring a KIT exon 11 mutation have a good response to TKI treatment, whereas the same treatment is less effective in tumours with KIT exon 9 mutations [13]. Additionally, progressive disease is common during long-term TKI treatment due to acquired treatment resistance [14,15].

In the era of personalized medicine, it is of utmost importance to evaluate the efficacy of TKI treatment at an early stage in order to optimize therapy strategies and protect patients from futile ineffective treatment, unnecessary side-effects and healthcare costs. Contrast-enhanced computed tomography (CE-CT) and 2-deoxy-2-[^{18}F]fluoro-D-glucose ([^{18}F]FDG) PET/CT are considered useful for diagnosis and response monitoring in GIST patients. The imaging modalities offer information on tumour morphology, perfusion characteristics, as well as tumour glucose metabolism [7]. However, optimal use of imaging for predicting and monitoring TKI treatment in patients with GIST is still a subject of debate. This systematic review aims to elucidate the added value of CE-CT and [^{18}F]FDG PET/CT imaging in the *prediction of response* and *early response monitoring* of TKI treatment in localized and advanced GISTs.

2. Methods

2.1 Search strategy

From 29 April 2022 to 24 June 2022, the databases of PubMed, Web of Science, the Cochrane Library and Embase were systematically screened using predefined search queries (Supplementary Materials). The following terms and their corresponding synonyms were included: “gastrointestinal stromal tumour”, “(neo)adjuvant”, “TKI treatment” and “FDG-PET” and “Tomography, X-ray Computed” imaging. The search queries are wide-ranging and seek to cover the aspect of both response monitoring and prediction by including ‘monitoring’ and ‘prediction models’ as well as ‘radiomics’ and ‘prognostics’. In addition to these search terms, other terms, such as ‘patient selection’ and ‘personalized medicine’, were also added, since these articles presumably covered the subject of TKI treatment evaluation and its efficacy in specific patient groups as well. The search strategy was implemented in consultation with an experienced research directorate, and access to the databases was granted by the Leiden University Medical Center.

2.2 Article selection

Articles were screened and considered eligible for full-text assessment if the title or abstract mentioned (i) quantitative outcome measures to evaluate the efficacy of imaging features (ii) retrieved from CE-CT and/or

[^{18}F]FDG PET/CT imaging (iii) in predicting or monitoring (neo)adjuvant TKI treatment response (iv) in localized and advanced GISTs. Response monitoring is defined as the evaluation of disease over the course of treatment using multiple medical imaging time points. Predicting response, however, solely involves the use of baseline scans made prior to TKI treatment administration. Articles assessing the prognostic value of different clinical and imaging parameters (e.g., risk of recurrence and metastatic potential) that can guide TKI treatment duration or timing for specific patient groups were also included, since these findings may improve patient selection in the future. Exclusion criteria comprised non-English and non-human studies, reviews, guidelines, recommendations, editorials, conference papers and abstracts. Case reports and studies analysing less than ten patients were also excluded. If the title and abstract did not contain sufficient information, full-text evaluation was used for judgement of relevance.

Subsequently, the articles were screened on full-text and excluded if they did not meet the previously mentioned inclusion criteria or if full-texts were not available. During this assessment, the focus was primarily on quantitative outcome measure(s) of studies. Outcome measures that were included in this analysis were correlations, associations, area under the curve (AUC), sensitivity, specificity and accuracy.

Finally, the reference lists from included articles were screened to find additional articles on this topic. The articles were independently assessed by the first two authors (Y.A.W., G.M.K.) and in cases of discrepancy, consensus reading was performed to make a final decision that led to either inclusion or exclusion.

2.3 Quality assessment

Articles using a radiomics pipeline were assessed through the radiomics quality score (RQS). The RQS is a scoring system that assigns points to a radiomics study based on specific criteria, where a maximum score of 36 points can be awarded. In this paper, the RQS is modified to focus on the methodological aspects of the included studies. The following criteria were omitted from the RQS, yielding a modified RQS (RQS_m); 'imaging at multiple time points', 'trial database registry' and 'multivariable analysis on non-radiomics features' , since they were considered less relevant for the quality of the obtained models [16]. The criteria from the RQS_m were also used to create a quality assessment tool to assess studies on non-radiomics prediction models and correlational research. Modifying the RQS_m for non-radiomics studies yielded the $RQS_{m,nonrad}$. This $RQS_{m,nonrad}$ had a maximum score of ten points (Supplementary Materials). Articles were considered high quality if they reached a score above 50%. To assess applicability concerns and the risk of bias in articles covering the topic of monitoring, the Quality Assessment of Diagnostic Accuracy Studies Tool, Version 2 (QUADAS-2) was applied [17]. Articles on response monitoring were considered to have a high risk of bias or applicability concerns if two or more of the domains were scored as 'high' or 'unclear'. Subsequently these articles were scored as low-quality.

Quality assessment was performed by the first author (Y.A.W.). The quality score was not considered as an exclusion criterion, as the authors considered it important to review all relevant evidence [17-19].

2.4 Data analysis

The eligible studies were categorized based on their topic concerning either response prediction or therapy monitoring. From each study, detailed information on the publication year, first author, patient groups, type of TKI treatment and imaging technique(s), was obtained. The specific CE-CT and [^{18}F]FDG-PET imaging features and their corresponding conclusions on efficacy, along with the attributed quality score, were briefly summarized. In the results section, only studies that were considered to be high-quality, were analysed in depth by clarifying conclusions on clinical relevance, discrepancies in results and insights on biological correlates.

2.5 Response prediction

In response prediction, imaging features from baseline/diagnostic CE-CT and [^{18}F]FDG-PET/CT are retrieved to predict responder status, prior to TKI administration. Articles on this topic were divided into five categories: mutational status, proliferative activity, risk stratification, radiological response and prognosis. These categories were considered important, as they all influence treatment strategies. Clinical genotyping is essential for clinical decision making, regarding neoadjuvant therapy, since the sensitivity and resistance towards TKI treatment in GISTs is dependent on the mutational status. In addition, patients with a high-risk GIST (and thus high proliferative activity) receive adjuvant TKI treatment for a period of three years to eliminate remaining disease and reduce chances of relapse [7]. Predicting whether patients will have a radiological response or a good prognosis at baseline could also aid towards a more personalized TKI treatment.

2.6 Therapy monitoring

In therapy monitoring, one uses the visual and quantitative differences between baseline and follow-up scans to determine treatment response. The efficacy of CE-CT and [^{18}F]FDG-PET are first discussed separately, followed by a qualitative comparison between both imaging modalities.

3 Results

3.1 Search strategy and article selection

The search query identified a total of 599 articles from the databases of PubMed, Web of Science, the Cochrane library and Embase. The study selection process led to a total of 90 articles eligible for analysis (Fig. 1). Articles that were excluded based on imaging criteria included, for example, the use of radiotracers other than [^{18}F]FDG [20]. Additionally, some articles discussed the use of molecular genotyping and DNA

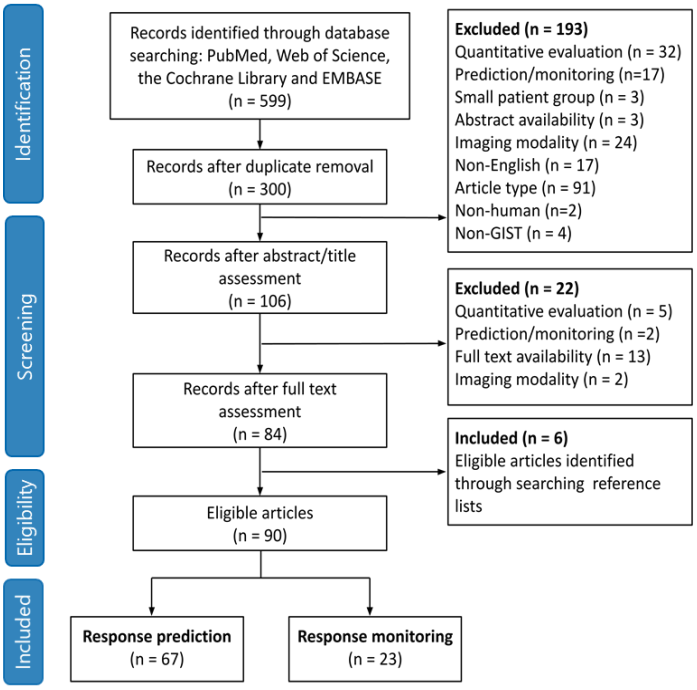

Fig. 1. Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) flowchart showing all the exclusion criteria. A total of 90 articles were included for this systematic review. Sixty-seven articles covered the topic of response prediction, and 23 articles covered the topic of response monitoring.

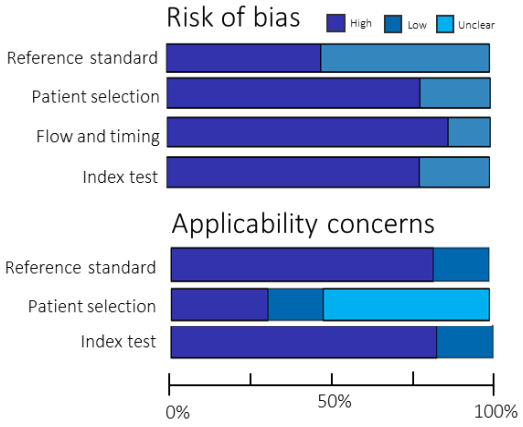

Fig. 2. Summary of Methodological Quality Scored According to Quality Assessment of Diagnostic Accuracy Studies Tool, Version 2 (QUADAS-2) for 23 articles discussing the topic of response monitoring.

sequencing to predict or determine response and therefore did not involve the use of any imaging modality [21]. Other excluded articles discussed the efficacy of a specific TKI treatment but did not quantify the efficacy of imaging features in predicting or monitoring response [22,23]. Of the 90 eligible articles, 67 were concerning response prediction [24-90] and 23 discussed response monitoring [91-113].

3.2 Quality assessment

Twenty-two articles discussed the use of radiomic models, and six out of 22 studies were of low quality (score < 50%). The mean RQS_m of the included articles was 13.5 (SD ± 2.60) out of 26 points. Low scores were mainly caused by a lack of transparency, biological correlates and gold standard comparison. Two articles received a score of 18 points (69.2%), which was the highest attributed score [70,88]. The forty-five studies on non-radiomic prediction models and correlational research scored an average $RQS_{m,nonrad}$ of 3.91 (SD ± 1.23) out of ten points, where eighteen articles scored above 50.0%. This was mainly caused by the fact that only a few articles used gold standard comparison [31,35,46] or an undescribed test set to validate their results [44,48,68]. The results of the QUADAS-2 tool are graphically displayed in Figure 2. Eight articles on response monitoring had high risk of bias or concerns for applicability and were therefore scored as low-quality. High risk of bias was often introduced by using reference standards involving follow-up (e.g., progression free survival, overall survival, time-to-treatment failure). Concerns for applicability were mainly caused by a lack of reporting on patient characteristics. In this way, judgement on whether the included patients matched the review question was unclear.

3.3 Response prediction

All articles on response prediction have been summarized in the Supplementary Materials. In this section only high-quality scored studies will be discussed.

3.3.1 Mutational status

The radiomic model of Starmans et al. was validated on unseen data and achieved an AUC, sensitivity and specificity of 0.51, 96.0% and 3.00% for predicting KIT mutation presence [81]. The model, based on portal venous radiomic features, requires further improvement in order to be clinically applicable.

Three studies developed a model or nomogram based on radiomic features obtained from CE-CT imaging (arterial, venous and delayed phase) to predict the presence of KIT exon 11 mutations, which resulted in varying AUC outcomes, namely 0.57, 0.72 and 0.81 [75,76,81]. Deletions in exon 11 may indicate more aggressive tumour behaviour, and for this reason, Liu et al. also assessed the efficacy of their model in predicting exon 11 deletion affecting codons 557–558 and achieved an AUC of 0.85 [76].

In clinical practice, patients with KIT exon 9 mutations often receive a high-dose imatinib regimen (800 mg) to improve progression-free survival (PFS). Yin et al. showed significantly greater tumour sizes and higher enhancement ratios (Hounsfield units (HU) for tumour parenchyma divided by the HUs of the erector spinae muscle) on portal venous CE-CT imaging compared to KIT exon 11 mutations. Using a 1.60 cut-off point, KIT exon 9 mutated small intestine tumours could be differentiated with an AUC, sensitivity and specificity of 0.76 and 86.7% and 98.5%, respectively. This threshold has, however, not been validated on independent validation data [67].

3.3.2 Proliferative activity

Since high-risk GISTs have a high proliferation rate, several studies attempted to link the mitotic index and Ki-67 proliferation index to imaging features in order to make a non-invasive assessment of expected tumour behaviour. On CE-CT imaging, intralesional hypodensity and concurrent heterogeneous enhancement patterns were significantly more common in high-mitotic tumours (Figures 3 and 4) [29,46].

Hypodensity was, in this case, defined as an area of low attenuation on portal venous phase CE-CT with Hus between 0 and 30 and when no HU increase (max 5 HUs) was observed between unenhanced and post-contrast images [46]. The changes in enhancement patterns were attributed to the principle of neovascularisation. Tumours with high proliferative activity can induce the formation of hyperpermeable disorganized blood vessels and consequent development of necrosis [29,61]. Therefore, the supply and washout of contrast agent is affected, which has a direct impact on tumour enhancement patterns.

A radiomic model using 42 quantitative and semantic imaging features (tumour location, first-order and texture radiomic features) retrieved from portal venous CE-CT imaging, differentiated high- from low-mitotic tumours with an AUC, sensitivity and specificity of 0.54, 27.0% and 75.0%, respectively [81]. Although on theoretical grounds CE-CT should be able to visualize poor neo-vasculature due to rapid tumour growth, no radiomic study has been able to establish this correlation. However, radiomic models predicting high Ki-67 proliferation index in localized and advanced GISTs achieved AUC values above 0.75 [77,88,89].

Comparison of studies investigating the relation between imaging and Ki-67 indices is complicated by the fact that different thresholds (e.g., 4%, 5%, 8% and 10%) for Ki-67 expression were used. Due to the small study sizes and heterogeneous outcomes with respect to Ki-67 indices, the true relationship between CE-CT imaging and proliferation has yet to be established.

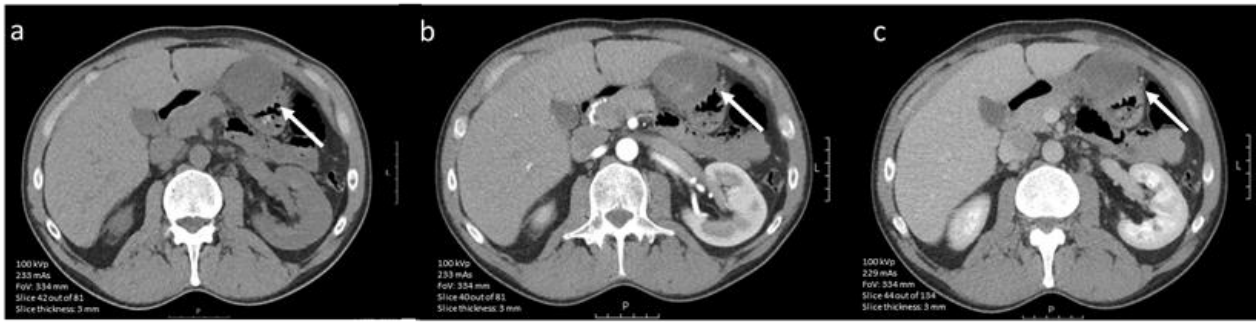

Fig. 3 (a-c). Axial contrast-enhanced (iodinated contrast media) CT image of a 45-year-old male diagnosed with a (histopathologically confirmed) low mitotic gastric GIST. The lesion (arrow) shows a round tumor shape and homogeneous enhancement in (**a**) nonenhanced phase, (**b**) arterial phase and (**c**) portal venous phase (scale bars 5 cm).

Fig. 4 (a-c). Axial contrast-enhanced (iodinated contrast media) CT images of a 60-year-old male diagnosed with a (histopathologically confirmed) high-mitotic gastric GIST. The lesion (arrow) shows a lobulated tumor shape, heterogeneous enhancement in (**a**) nonenhanced phase, (**b**) arterial phase and (**c**) portal venous phase (scale bars 5 cm).

3.3.3 Risk stratification

Research on the use of [^{18}F]FDG-PET imaging features for risk stratification in GISTs is limited. In two studies, high metabolic tumour volume (MTV) and total lesion glycolysis (TLG) were predictive for high risk GISTs [25,35]. The use of quantitative imaging features showed improved predictive accuracy during follow-up when compared to a clinical reference standard (NIH modified criteria) [35]. Although these results suggest

the added role of [^{18}F]FDG-PET for risk stratification, there are only a few studies that investigated [^{18}F]FDG-PET for this purpose.

Larger tumour sizes, mixed or extra-luminal growth patterns, ill-defined tumour shape, presence of vessels feeding or vessels draining the tumour mass, necrosis and ulceration on CE-CT imaging were all associated with high-risk GISTs [44,53,58,60,63,64,68]. Of note, Wei et al. used the angle between the longest and shortest tumour diameter to quantify tumour shape. This parameter was able distinguish intermediate- and high-risk from low-risk GISTs more accurately when compared to using solely the longest diameter [58]. Heterogeneous enhancement patterns on portal venous phase CE-CT proved to be predictive for high-risk GISTs as well (Figure 5) [53]. Incomplete enhancement of the overlying gastric mucosa on arterial phase, was also significantly more common in high-risk gastric GISTs [51]. In a study by Tang et al., HUs of the arterial phase CE-CT were subtracted from the attenuation coefficients in the portal venous phase to derive quantitative features describing contrast enhancement. Using the subtraction CT, small regions of interest (ROIs) of 30–50 mm2, were placed in the most enhancing solid components of the tumours. The difference in HUs was significantly lower in high-risk gastric GISTs [53]. Additionally, the peak value of enhancement on CE-CT (arterial and portal venous phase) imaging was strongly correlated with risk [45]. Both articles suggest a rapid inflow of iodinated contrast agent in high-risk GISTs and thus the presence of permeable and leaky tumour vessels. The mean of the positive pixels (HU > 0) of the entire tumour volume on portal venous CT imaging was lower in high-risk GISTs [31]. This observation can be attributed to the presence of tumour necrosis, which was more commonly found in the high-risk group.

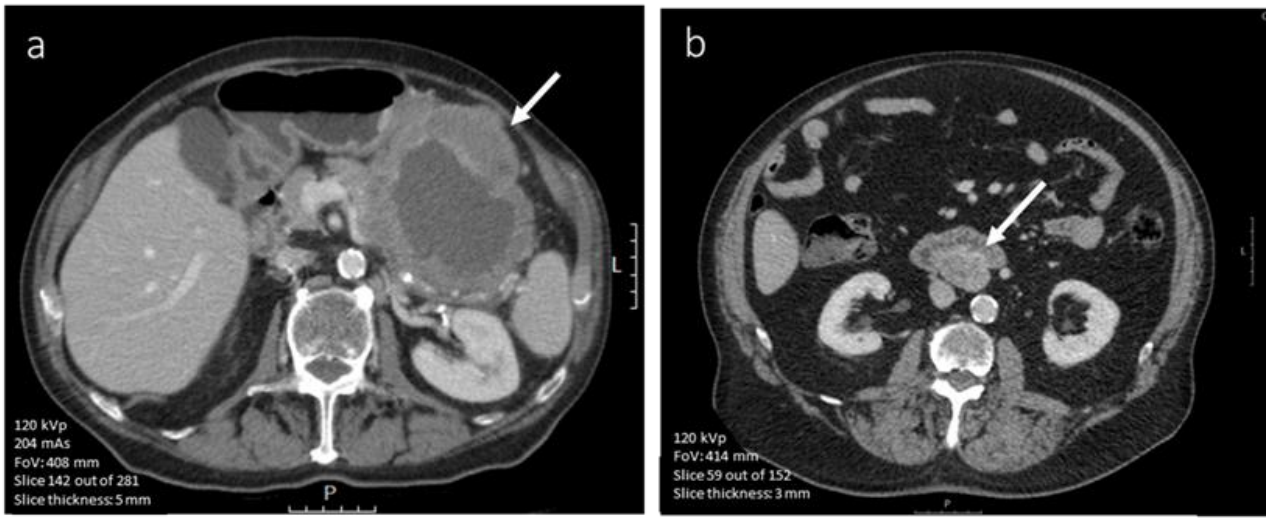

Fig. 5 (a-b) Axial portal venous phase (iodinated contrast media) CT image of an 83-year-old male diagnosed with a high-risk (Miettinen AFIP classification) gastric GIST (scale bar 5 cm). The large lesion is lobulated and has central necrosis (arrow). (**b**) Axial portal venous phase (iodinated contrast media) CT slice of a 72-year-old male diagnosed with a low-risk GIST affecting the small intestine (scale bar 5 cm). It shows a well-defined and rounded lesion with a homogeneous enhancement pattern (arrow).

By contrast, Li et al. included gastric, intestinal and extra gastrointestinal tumours and did not find a significant difference in enhancement patterns between risk groups [43]. Although tumour enhancement has been established as a relevant factor in the risk stratification of GISTs, there are discrepancies in the results.

Machine learning used for the prediction of risk is extensively investigated with a total of twelve articles covering this topic [71,72,79,83,86,87]. All models achieved an AUC above 0.83 for predicting high-risk GISTs, with an average AUC of 0.87. In many of the models, texture radiomic features (grey level co-occurrence matrix (GLCM), neighbouring grey-tone difference matrix (NGTDM) and grey run-length matrix (GRLM) and

grey level size zoned matrix (GLSZM)) were used to develop the model. These texture features reflect enhancement patterns and inter-pixel relationships in a three-dimensional tumour volume.

3.3.4 Prediction of radiological response

There was one article attempting to predict radiological response using baseline imaging. Disease progression was in this case defined by the modified Choi criteria, which is currently one of the reference standards used for GISTs response evaluation [114]. In this case, the Choi criteria were combined with four textural portal venous features (features retrieved from GLCM, GLRLM and NGTDM), disease progression was predicted with an AUC of 0.827 [32].

3.3.5 Prognosis

Of the selected articles, two articles discussed the use of imaging features obtained from [^{18}F]FDG-PET/CT imaging to predict progression free survival (PFS), through detection of disease recurrence (locally and or development of distant metastases). They found significantly higher MTV and TLG values in patients with a lower PFS. In addition to quantitative [^{18}F]FDG-PET imaging features, larger tumour sizes were also a significant factor contributing to lower PFS [25,35].

On CE-CT imaging, one study with a relatively large patient group (n = 143), observed that tumour sizes greater than 10 cm, ill-defined tumour outline and enhancing solid components contributed to a poor patient prognosis, as reflected by their OS [48]. The study by Jung et al. combined relevant predictive parameters (tumour location, ill-defined tumour outline and presence of feeding vessels) to create a nomogram. The nomogram was internally validated and achieved an AUC of 0.863 [37]. In addition to semantic CT features, Ekert et al. assessed the efficacy of four quantitative textural features (GLCM inverse difference normalized, GLRLM normalized, GLRLM normalized and NGTDM coarseness) from portal venous phase CT imaging to predict prognosis of GIST patients. This study found that high values for these texture features were all associated with poor PFS [32].

In another study, three-year recurrence free survival (RFS) was predicted by a deep learning ResNet model based on features retrieved from arterial phase images. Results show that, using an internal validation cohort, a predictive model with an AUC of 0.912 was obtained [70]. Furthermore, Zheng et al. investigated whether the occurrence of liver metastasis in high risk GISTs could be predicted. They found that a model based on portal venous CT radiomic features reached an AUC and accuracy of 0.873 and 84.9% [90].

3.4 Therapy monitoring

All articles on therapy monitoring have been summarized in the Supplementary Materials. In this section only high-quality scored studies will be discussed.

3.4.1. CE-CT Imaging

Many articles discussed the use of the well-established Response Evaluation Criteria in Solid Tumours (RECIST 1.1) to assess tumour response. RECIST 1.1 is a method in which the sum of the longest diameter of (a maximum of 5) target lesions is used to evaluate treatment response. The RECIST 1.1 scoring system categorizes patients into four types of response, namely complete response (disappearance of all lesions), partial response (≥30% reduction of the sum of the target lesions (SLD)), progressive disease (≥20% increase of the SLD compared to the smallest SLD ever measured) and stable disease (neither progressive disease nor partial response) [115]. Nonetheless, substantial tumour shrinkage is often not observed during effective TKI treatment. Subtle and moderate changes in tumour size may be more accurately quantified by means of volumetric measurements. This is shown by Schiavon et al., who showed that size changes in GIST liver metastases larger than 20% were more frequently detected by volumetric measurements compared

to the RECIST 1.1 criteria [110]. By using solely one-dimensional measurements, one presumes tumours remain spherical and that response occurs equally along three orthogonal axes during TKI treatment. However, liver metastasis in GIST patients showed significant changes in morphology over the course of imatinib treatment, which was better reflected by an ellipsoid volumetric approach [109].

In addition to RECIST 1.1, Choi et al. proposed a new method (Choi criteria) by including treatment-related changes in portal venous CT tumour densities [95]. Suppression of vascular endothelial growth factor expression can be induced by TKI treatment [116,117]. Therefore, treatment leads to changes in tumour vascularity and can lead to a reduction in tumour density, as reflected by the value of the HUs measured on CT (Fig. 6). Using RECIST 1.1 and Choi, comparable results were obtained for predicting PFS for patients treated with second line sunitinib assessed during an early follow-up of about 2–3 months [96,105]. Nonetheless, the Choi criteria gradually overestimated the number of patients with a partial response to sunitinib and regorafenib during longer follow-up periods (up to a year), leading to poorer PFS [105,106]. It was speculated that a drop in tumour density could also be caused by tumour necrosis, which is often a sign of progressive disease. So, instead of measuring a reduction in tumour vascularization, one may be measuring progressive disease over longer follow-up periods [105].

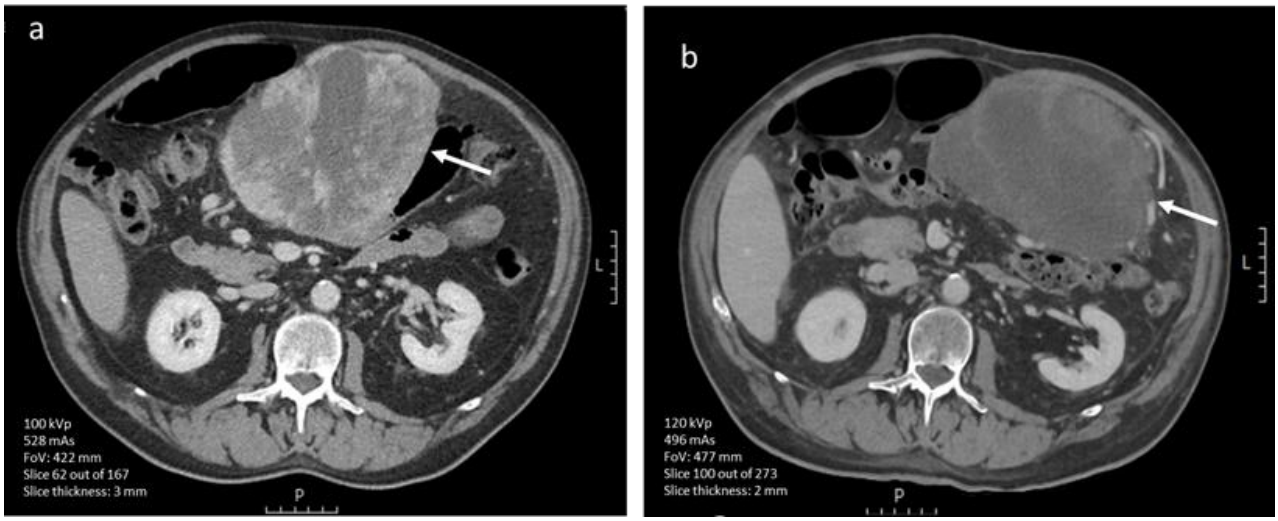

Fig. 6 (a-b). Axial portal venous phase CT images (iodinated contrast media) of a 67-year-old male diagnosed with a primary GIST (arrows) of the stomach. (**a**) Pre-treatment imaging shows a large gastric mass with heterogeneous enhancement. (**b**) After 1.5 months of avapritinib treatment, the lesion has become hypodense (scale bars 5 cm).

3.4.2 [^{18}F]FDG-PET imaging

In [^{18}F]FDG-PET imaging, the European Organization for Research and Treatment in Cancer (EORTC) PET criteria are most commonly used, in which a metabolic response is determined by a reduction in SUVmax of 25% or more [118]. Metabolic response was significantly associated with prolonged PFS and could be detected as early as seven days, after the induction of TKI treatment (imatinib and sunitinib) [100,102]. On the contrary, the prospective study of Chacón et al. did not find a significant association between PFS and metabolic response determined by the EORTC PET criteria.

Additionally, two retrospective studies by Farag et al. evaluated the impact of [^{18}F]FDG-PET/CT on clinical decision making in the treatment of localized and advanced GIST patients. Changes in surgical management, systemic treatment and treatment objective were all included in the evaluation [111,112]. In 27.1% of GIST patients treated with neoadjuvant intent, management was changed because of [^{18}F]FDG-PET/CT findings at an interval of eight weeks. The lack of metabolic response was correlated with therapeutic changes in management, especially in non-KIT exon 11 mutations [111]. In the advanced disease setting, specifically

late [^{18}F]FDG-PET response findings (median of 293 days) proved to have an impact on therapeutic decision [112].

3.4.3 CE-CT vs. [^{18}F]FDG-PET imaging

When comparing the aforementioned response evaluation criteria on CE-CT imaging with the EORTC PET criteria on [^{18}F]FDG-PET imaging, articles reported high agreement and RECIST responders also showed significant reductions in SUV_{max} [91,98,100,108]. Choi et al. showed greater sensitivity and specificity (97.0% and 100%) when compared to the EORTC PET criteria [95]. Metabolic response could, however, be observed within a week and preceded changes in tumour size and volume in localized and advanced GIST patients treated with imatinib (Fig. 7) [92,97,100,107]. By using the RECIST 1.1 criteria, the early effect of TKI treatment may be underestimated. For example, Choi et al. showed that 70% of the stable disease RECIST patients had a SUV_{max} reduction between 61 and 100% at two months follow-up [94].

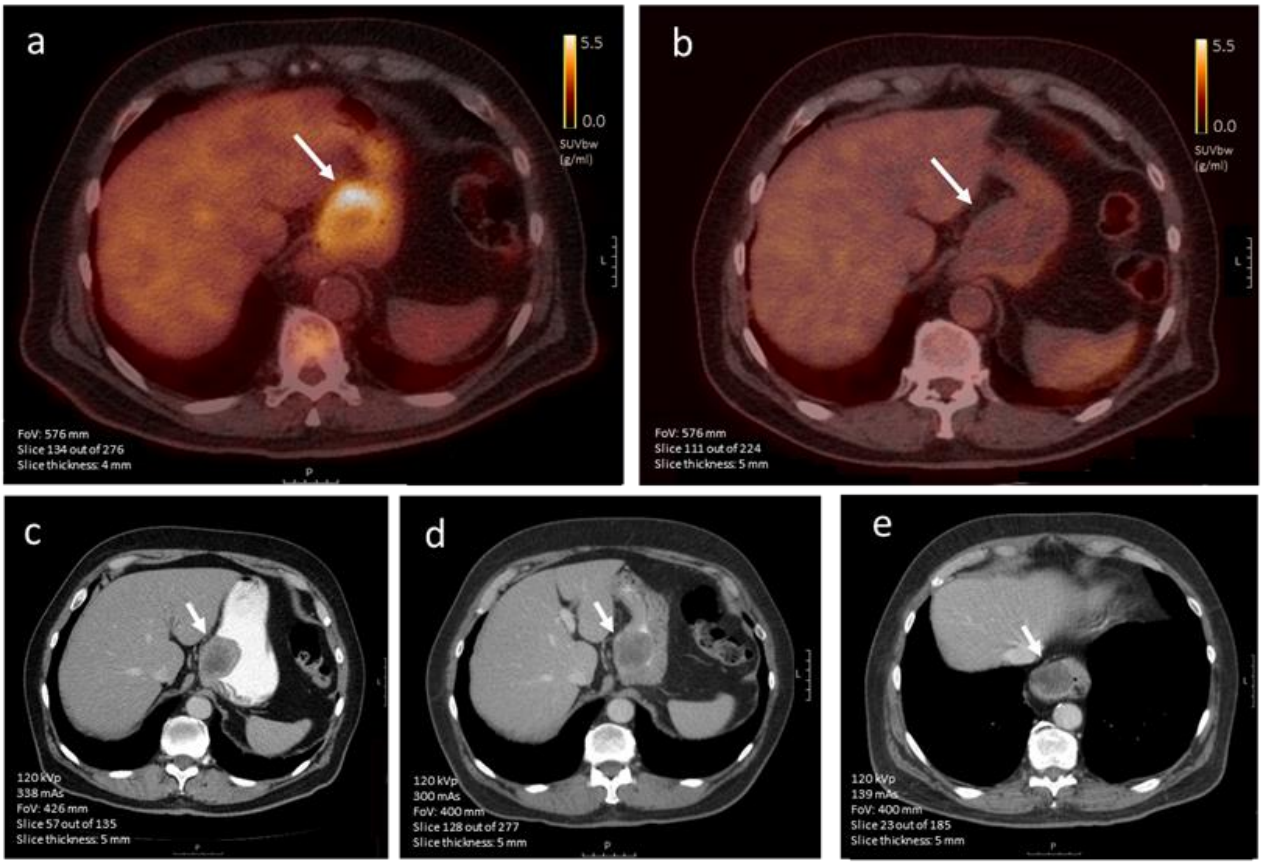

Fig. 7 (a-e). Axial [^{18}F]FDG-PET images of a 71-year-old male diagnosed with a primary gastric GIST a) before treatment induction and b) after about three months of TKI treatment, where the standardized uptake value (SUV) is normalized. Corresponding contrast-enhanced CT imaging visualizing the same lesion (arrow) c) at diagnosis and after d) 2.5 months and e) 6.5 months of imatinib treatment, showing minimal to no change in tumor size. In the last image, the intrathoracic tumor location is caused by a sliding hernia.

4 Discussion

The aim of this review was to provide an overview of the value of CE-CT and [^{18}F]FDG PET/CT imaging to predict and monitor TKI treatment response in GIST patients.

There is limited literature available on the use of baseline [^{18}F]FDG-PET findings to predict tumour response. Although there are only a few studies available, generally imaging features, such as MTV and TLG, were correlated with more aggressive tumour behaviour. On the contrary, there is more data available on the potential of CE-CT imaging features to predict treatment response. Results indicate that larger tumour sizes

(>5 cm), ill-defined or lobulated tumour outline, mixed or exophytic growth patterns, the presence of (enlarged) and feeding vessels are associated with patient outcome. The presence of heterogeneous enhancement patterns was a recurring observation in high-risk GISTs. The hypodensities observed on CE-CT imaging were devoted to the biological phenomena of neovascularisation and necrosis. It should be noted that the correlation between hypodensities on radiological imaging and actual pathological necrosis and neovascularisation in GIST tumours is still disputable.

Many articles discussed the use of radiomic and deep learning models for response prediction on baseline CE-CT imaging. High performance scores were stated for models predicting RFS and risk stratifications, while mutational status remained difficult to predict with variable AUC values. Radiomics offers the possibility to identify clinically relevant imaging features that would normally be imperceptible to the naked eye. For example, it has proven to be difficult to obtain a sufficient amount of tissue samples from biopsy material, which makes it difficult to determine the mutational status or a reliable mitotic count. Additionally, if the mitotic count is determined on postoperative surgical specimens, the results can be inaccurate due to the occasional administration of neoadjuvant TKI treatment. It would, therefore, be very helpful if imaging could provide additional information, other than tumour size. Nonetheless, the biological explanation behind the efficacy of radiomic features was often missing in the included articles. Before advanced and objective learning techniques can be introduced in clinical practice, they should be clinically relevant and biologically meaningful. It is recommended to further explore the prediction of actual radiological response using semantic or quantitative imaging features selected based upon tumour biology.

The three evaluation methods currently used to monitor response in GIST patients, are the RECIST 1.1, Choi and EORTC PET criteria. The main disadvantage of the RECIST 1.1 criteria is the one-dimensional nature of its measurements, presuming a spherical tumour shape throughout the entire course of TKI treatment. To overcome this limitation, an additional set of criteria was developed by Choi et al. involving CT densities. The Choi criteria are occasionally applied in clinical practice. However, its efficacy and prognostic value in determining response in GIST patients remains unclear. Supposedly, the antiangiogenic effect of TKI treatment would lead to a consequent reduction in HU values. As previously stated, necrosis and heterogeneous enhancement patterns at baseline were considered predictive for more aggressively behaving tumours. Using reductions in CT densities as a criterion for response monitoring may, therefore, be misleading, since it can reflect a decrease in angiogenesis induced by TKI treatment, as well as necrosis induced by aggressive tumour behaviour. This hypothesis was supported by literature, since response evaluation using Choi criteria led to an overestimation in the number of partial responders at longer follow-up periods.

[^{18}F]FDG-PET proved to be useful in the early monitoring of GISTs, since significant reductions in SUVmax could be observed within a week of TKI treatment and metabolic changes preceded morphological changes in size. However, this imaging technique is often not considered for early response monitoring in clinical practice because of higher costs. Since some of the targeted treatments are more expensive than PET-CT scans, further research should, therefore, be focused on the cost-effectiveness of [^{18}F]FDG-PET imaging in the treatment of GISTs.

Particularly, the combined use of different imaging modalities, also known as multimodality imaging, might provide more detailed information that can assist in making early image-guided treatment decisions. The use of such a multimodality imaging approach might be useful to gather as much information as possible on the biological behaviour of GIST. However, currently, no literature is available on the specific use of combining these different imaging modalities for response prediction or monitoring.

5 Conclusion

In conclusion, imaging features obtained from CE-CT and [^{18}F]FDG PET/CT imaging can aid towards a more personalized treatment of GIST patients by enabling early prediction and monitoring of TKI therapy response. Heterogeneous enhancement patterns on baseline CE-CT imaging were predictive for high risk GISTs, reflecting neovascularisation and necrosis.

For the purpose of response monitoring, current RECIST 1.1 and Choi criteria are still lacking sensitivity or prone to errors when predicting or monitoring treatment response. [^{18}F]FDG-PET is a promising imaging technique that visualizes functional metabolic changes in GISTs, which precedes measurable changes in tumour size. Although promising, the true added value of [^{18}F]FDG-PET remains elusive and research on cost-effectiveness is warranted.

Radiomics is an emerging topic in medicine and shows potential for the prediction of RFS and risk stratifications in GISTs. However, future research should mainly focus on clinical utility, explainability, and correlation with actual tumour biology.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/diagnostics12112722/s1.

Author contributions: Conceptualization, Y.A.W., G.M.K., J.A.v.d.H., W.G. and L.-F.d.G.-O.; methodology, Y.A.W., G.M.K., W.G., J.A.v.d.H. and L.-F.d.G.-O.; software; Y.A.W. and G.M.K.; validation, Y.A.W., G.M.K. and W.G.; formal analysis, Y.A.W., G.M.K., W.G. and L.-F.d.G.-O.; investigation, Y.A.W., G.M.K., W.G. and L.-F.d.G.-O.; resources, Y.A.W., G.M.K., W.G., J.A.v.d.H. and L.-F.d.G.-O.; data curation, Y.A.W., G.M.K., W.G. and L.-F.d.G.-O.; writing—Original draft preparation, Y.A.W., G.M.K., W.G. and L.-F.d.G.-O.; writing—Review and editing, G.M.K., W.G., J.A.v.d.H., F.H.P.v.V., H.G., J.V.M.G.B., A.J.v.d.M. and L.-F.d.G.-O.; visualization, Y.A.W., A.J.v.d.M. and L.- F.d.G.-O.; supervision, G.M.K., W.G. and L.-F.d.G.-O.; project administration, Y.A.W. All authors have read and agreed to the published version of the manuscript.

Funding: We declare the following financial interests/personal relationships which may be considered as potential competing interests: G.M. Kalisvaart was the recipient of an educational grant (LEI-05) from Philips Electronics Nederland B.V, Eindhoven and supported by a public grant (LSHM18089) from Health Holland TKI Life Sciences & Health. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data, in the writing of the manuscript or in the decision to publish the results.

Institutional Review board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee Leiden Den Haag Delft (METC LDD) (protocol code: B19.050, date of approval: 14 January 2020).

Informed Consent Statement: Patient consent was waived due to the retrospective nature of the study. Patients who objected to the use of their data were excluded.

Data Availability Statement: The datasets generated during and/or analysed during the current study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors would like to thank Jan W. Schoones for his contribution to the search strategy.

Conflict of interest: The other authors declare no conflict of interest.

6 References

1. Balachandran, V.P.; DeMatteo, R.P. Gastrointestinal stromal tumors: who should get imatinib and for how long? *Adv Surg* 2014, *48*, 165-183, doi:10.1016/j.yasu.2014.05.014.
2. Zappacosta, R., Zappacosta, B., Capanna, S., D'Angelo, C., Gatta, D., Rosini, S. . GISTs: From the History to the Tailored Therapy, Gastrointestinal Stromal Tumor. *IntechOpen* 2012, doi:10.5772/33925.
3. Joensuu, H.; Hohenberger, P.; Corless, C.L. Gastrointestinal stromal tumour. *Lancet* 2013, *382*, 973-983, doi:10.1016/s0140-6736(13)60106-3.
4. Wu, C.E.; Tzen, C.Y.; Wang, S.Y.; Yeh, C.N. Clinical Diagnosis of Gastrointestinal Stromal Tumor (GIST): From the Molecular Genetic Point of View. *Cancers (Basel)* 2019, *11*, doi:10.3390/cancers11050679.
5. Verweij, J.; Casali, P.G.; Zalcberg, J.; LeCesne, A.; Reichardt, P.; Blay, J.Y.; Issels, R.; van Oosterom, A.; Hogendoorn, P.C.; Van Glabbeke, M.; et al. Progression-free survival in gastrointestinal stromal tumours with high-dose imatinib: randomised trial. *Lancet* 2004, *364*, 1127-1134, doi:10.1016/s0140-6736(04)17098-0.
6. Reichardt, P. The Story of Imatinib in GIST - a Journey through the Development of a Targeted Therapy. *Oncol Res Treat* 2018, *41*, 472-477, doi:10.1159/000487511.
7. Casali, P.G.; Blay, J.Y.; Abecassis, N.; Bajpai, J.; Bauer, S.; Biagini, R.; Bielack, S.; Bonvalot, S.; Boukovinas, I.; Bovee, J.; et al. Gastrointestinal stromal tumours: ESMO-EURACAN-GENTURIS Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2022, *33*, 20-33, doi:10.1016/j.annonc.2021.09.005.
8. Casali, P.G.; Le Cesne, A.; Poveda Velasco, A.; Kotasek, D.; Rutkowski, P.; Hohenberger, P.; Fumagalli, E.; Judson, I.R.; Italiano, A.; Gelderblom, H.; et al. Time to Definitive Failure to the First Tyrosine Kinase Inhibitor in Localized GI Stromal Tumors Treated With Imatinib As an Adjuvant: A European Organisation for Research and Treatment of Cancer Soft Tissue and Bone Sarcoma Group Intergroup Randomized Trial in Collaboration With the Australasian Gastro-Intestinal Trials Group, UNICANCER, French Sarcoma Group, Italian Sarcoma Group, and Spanish Group for Research on Sarcomas. *J Clin Oncol* 2015, *33*, 4276-4283, doi:10.1200/jco.2015.62.4304.
9. Lopes, L.F.; Bacchi, C.E. Imatinib treatment for gastrointestinal stromal tumour (GIST). *J Cell Mol Med* 2010, *14*, 42-50, doi:10.1111/j.1582-4934.2009.00983.x.
10. Reynoso, D.; Trent, J.C. Neoadjuvant and adjuvant imatinib treatment in gastrointestinal stromal tumor: current status and recent developments. *Curr Opin Oncol* 2010, *22*, 330-335, doi:10.1097/CCO.0b013e32833aaaad.
11. Eisenberg, B.L.; Trent, J.C. Adjuvant and neoadjuvant imatinib therapy: current role in the management of gastrointestinal stromal tumors. *Int J Cancer* 2011, *129*, 2533-2542, doi:10.1002/ijc.26234.
12. Yang, W.; Liu, Q.; Lin, G.; Zhang, B.; Cao, H.; Zhao, Y.; Xia, L.; Feng, F.; Xiong, Z.; Hu, J.; et al. The effect of neoadjuvant imatinib therapy on outcome and survival in rectal gastrointestinal stromal tumors: A multiinstitutional study. *J Surg Oncol* 2021, *124*, 1128-1135, doi:10.1002/jso.26628.
13. Miettinen, M.; Lasota, J. Gastrointestinal stromal tumors: review on morphology, molecular pathology, prognosis, and differential diagnosis. *Arch Pathol Lab Med* 2006, *130*, 1466-1478, doi:10.5858/2006-130-1466-gstrom.
14. Betz, M.; Kopp, H.G.; Spira, D.; Claussen, C.D.; Horger, M. The benefit of using CT-perfusion imaging for reliable response monitoring in patients with gastrointestinal stromal tumor (GIST) undergoing treatment with novel targeted agents. *Acta Radiol* 2013, *54*, 711-721, doi:10.1177/0284185113484642.
15. Cassier, P.A.; Fumagalli, E.; Rutkowski, P.; Schöffski, P.; Van Glabbeke, M.; Debiec-Rychter, M.; Emile, J.F.; Duffaud, F.; Martin-Broto, J.; Landi, B.; et al. Outcome of patients with platelet-derived growth factor receptor alpha-mutated gastrointestinal stromal tumors in the tyrosine kinase inhibitor era. *Clin Cancer Res* 2012, *18*, 4458-4464, doi:10.1158/1078-0432.Ccr-11-3025.
16. Lambin, P.; Leijenaar, R.T.H.; Deist, T.M.; Peerlings, J.; de Jong, E.E.C.; van Timmeren, J.; Sanduleanu, S.; Larue, R.; Even, A.J.G.; Jochems, A.; et al. Radiomics: the bridge between medical

imaging and personalized medicine. *Nat Rev Clin Oncol* 2017, *14*, 749-762, doi:10.1038/nrclinonc.2017.141.
17. Whiting, P.F.; Rutjes, A.W.; Westwood, M.E.; Mallett, S.; Deeks, J.J.; Reitsma, J.B.; Leeflang, M.M.; Sterne, J.A.; Bossuyt, P.M. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011, *155*, 529-536, doi:10.7326/0003-4819-155-8-201110180-00009.
18. Jüni, P.; Witschi, A.; Bloch, R.; Egger, M. The Hazards of Scoring the Quality of Clinical Trials for Meta-analysis. *JAMA* 1999, *282*, 1054-1060, doi:10.1001/jama.282.11.1054.
19. Whiting, P.; Harbord, R.; Kleijnen, J. No role for quality scores in systematic reviews of diagnostic accuracy studies. *BMC Med Res Methodol* 2005, *5*, 19, doi:10.1186/1471-2288-5-19.
20. Pretze, M.; Reffert, L.; Diehl, S.; Schönberg, S.O.; Wängler, C.; Hohenberger, P.; Wängler, B. GMP-compliant production of [(68)Ga]Ga-NeoB for positron emission tomography imaging of patients with gastrointestinal stromal tumor. *EJNMMI Radiopharm Chem* 2021, *6*, 22, doi:10.1186/s41181-021-00137-w.
21. Hedenström, P.; Andersson, C.; Sjövall, H.; Enlund, F.; Nilsson, O.; Nilsson, B.; Sadik, R. Pretreatment Tumor DNA Sequencing of KIT and PDGFRA in Endosonography-Guided Biopsies Optimizes the Preoperative Management of Gastrointestinal Stromal Tumors. *Mol Diagn Ther* 2020, *24*, 201-214, doi:10.1007/s40291-020-00451-0.
22. Cai, P.Q.; Lv, X.F.; Tian, L.; Luo, Z.P.; Mitteer, R.A., Jr.; Fan, Y.; Wu, Y.P. CT Characterization of Duodenal Gastrointestinal Stromal Tumors. *AJR Am J Roentgenol* 2015, *204*, 988-993, doi:10.2214/ajr.14.12870.
23. Nannini, M.; Pantaleo, M.A.; Maleddu, A.; Saponara, M.; Mandrioli, A.; Lolli, C.; Pallotti, M.C.; Gatto, L.; Santini, D.; Paterini, P.; et al. Duration of adjuvant treatment following radical resection of metastases from gastrointestinal stromal tumours. *Oncol Lett* 2012, *3*, 677-681, doi:10.3892/ol.2011.537.
24. Al-Balas, H.A.; Shaib, Y.H. Gastrointestinal stromal tumours: Role of computed tomography in predicting tumour behaviour. *Hong Kong Journal of Radiology* 2012, *15*, 155-161.
25. Albano, D.; Bosio, G.; Tomasini, D.; Bonù, M.; Giubbini, R.; Bertagna, F. Metabolic behavior and prognostic role of pretreatment 18F-FDG PET/CT in gist. *Asia Pac J Clin Oncol* 2020, *16*, e207-e215, doi:10.1111/ajco.13366.
26. Cannella, R.; Tabone, E.; Porrello, G.; Cappello, G.; Gozzo, C.; Incorvaia, L.; Grignani, G.; Merlini, A.; D'Ambrosio, L.; Badalamenti, G.; et al. Assessment of morphological CT imaging features for the prediction of risk stratification, mutations, and prognosis of gastrointestinal stromal tumors. *Eur Radiol* 2021, *31*, 8554-8564, doi:10.1007/s00330-021-07961-3.
27. Chen, T.; Xu, L.; Dong, X.; Li, Y.; Yu, J.; Xiong, W.; Li, G. The roles of CT and EUS in the preoperative evaluation of gastric gastrointestinal stromal tumors larger than 2 cm. *Eur Radiol* 2019, *29*, 2481-2489, doi:10.1007/s00330-018-5945-6.
28. Chen, X.S.; Shan, Y.C.; Dong, S.Y.; Wang, W.T.; Yang, Y.T.; Liu, L.H.; Xu, Z.H.; Zeng, M.S.; Rao, S.X. Utility of preoperative computed tomography features in predicting the Ki-67 labeling index of gastric gastrointestinal stromal tumors. *Eur J Radiol* 2021, *142*, 109840, doi:10.1016/j.ejrad.2021.109840.
29. Chen, Z.; Yang, J.; Sun, J.; Wang, P. Gastric gastrointestinal stromal tumours (2-5 cm): Correlation of CT features with malignancy and differential diagnosis. *Eur J Radiol* 2020, *123*, 108783, doi:10.1016/j.ejrad.2019.108783.
30. Cho, M.H.; Park, C.K.; Park, M.; Kim, W.K.; Cho, A.; Kim, H. Clinicopathologic Features and Molecular Characteristics of Glucose Metabolism Contributing to ^{18}F-fluorodeoxyglucose Uptake in Gastrointestinal Stromal Tumors. *PLoS One* 2015, *10*, e0141413, doi:10.1371/journal.pone.0141413.
31. Choi, I.Y.; Yeom, S.K.; Cha, J.; Cha, S.H.; Lee, S.H.; Chung, H.H.; Lee, C.M.; Choi, J. Feasibility of using computed tomography texture analysis parameters as imaging biomarkers for predicting risk grade of gastrointestinal stromal tumors: comparison with visual inspection. *Abdom Radiol (NY)* 2019, *44*, 2346-2356, doi:10.1007/s00261-019-01995-4.

32. Ekert, K.; Hinterleitner, C.; Horger, M. Prognosis assessment in metastatic gastrointestinal stromal tumors treated with tyrosine kinase inhibitors based on CT-texture analysis. *Eur J Radiol* 2019, *116*, 98-105, doi:10.1016/j.ejrad.2019.04.018.
33. Fuster, D.; Ayuso, J.R.; Poveda, A.; Cubedo, R.; Casado, A.; Martinez-Trufero, J.; Lopez-Pousa, A.; Del Muro, X.G.; Lomena, F.; Maurel, J.; et al. Value of FDG-PET for monitoring treatment response in patients with advanced GIST refractory to high-dose imatinib. A multicenter GEIS study. *Quarterly Journal of Nuclear Medicine and Molecular Imaging* 2011, *55*, 680-687.
34. Grazzini, G.; Guerri, S.; Cozzi, D.; Danti, G.; Gasperoni, S.; Pradella, S.; Miele, V. Gastrointestinal stromal tumors: relationship between preoperative CT features and pathologic risk stratification. *Tumori* 2021, *107*, 556-563, doi:10.1177/0300891621996447.
35. Hwang, S.H.; Jung, M.; Jeong, Y.H.; Jo, K.; Kim, S.; Wang, J.; Cho, A. Prognostic value of metabolic tumor volume and total lesion glycolysis on preoperative (18)F-FDG PET/CT in patients with localized primary gastrointestinal stromal tumors. *Cancer Metab* 2021, *9*, 8, doi:10.1186/s40170-021-00244-x.
36. Iannicelli, E.; Carbonetti, F.; Federici, G.F.; Martini, I.; Caterino, S.; Pilozzi, E.; Panzuto, F.; Briani, C.; David, V. Evaluation of the Relationships Between Computed Tomography Features, Pathological Findings, and Prognostic Risk Assessment in Gastrointestinal Stromal Tumors. *J Comput Assist Tomogr* 2017, *41*, 271-278, doi:10.1097/rct.0000000000000499.
37. Jung, H.; Lee, S.M.; Kim, Y.C.; Byun, J.; Park, J.Y.; Oh, B.Y.; Kwon, M.J.; Kim, J. Gastrointestinal stromal tumours: Preoperative imaging features to predict recurrence after curative resection. *European Journal of Radiology* 2022, *149*, 110193, doi:https://dx.doi.org/10.1016/j.ejrad.2022.110193.
38. Kim, H.C.; Lee, J.M.; Kim, K.W.; Park, S.H.; Kim, S.H.; Lee, J.Y.; Han, J.K.; Choi, B.I. Gastrointestinal stromal tumors of the stomach: CT findings and prediction of malignancy. *AJR Am J Roentgenol* 2004, *183*, 893-898.
39. Kamiyama, Y.; Aihara, R.; Nakabayashi, T.; Mochiki, E.; Asao, T.; Kuwano, H.; Oriuchi, N.; Endo, K. 18F-fluorodeoxyglucose positron emission tomography: useful technique for predicting malignant potential of gastrointestinal stromal tumors. *World J Surg* 2005, *29*, 1429-1435, doi:10.1007/s00268-005-0045-6.
40. Kim, H.C.; Lee, J.M.; Kim, S.H.; Park, S.H.; Lee, J.W.; Lee, M.; Han, J.K.; Choi, B.I. Small gastrointestinal stromal tumours with focal areas of low attenuation on CT: pathological correlation. *Clin Radiol* 2005, *60*, 384-388, doi:10.1016/j.crad.2004.06.022.
41. Kurata, Y.; Hayano, K.; Ohira, G.; Narushima, K.; Aoyagi, T.; Matsubara, H. Fractal analysis of contrast-enhanced CT images for preoperative prediction of malignant potential of gastrointestinal stromal tumor. *Abdom Radiol (NY)* 2018, *43*, 2659-2664, doi:10.1007/s00261-018-1526-z.
42. Kwon, Y.; Park, E.; Pahk, K.; Kim, S.; Kim, M.J.; Graf, D.; Park, S. Preoperative assessment of malignant potential of gastrointestinal stromal tumor by dual-time-point 18F-fluorodeoxyglucose positron emission tomography imaging: Usefulness of standardized uptake value and retention index. *J Cancer Res Ther* 2019, *15*, 142-147, doi:10.4103/jcrt.JCRT_1093_16.
43. Li, H.; Ren, G.; Cai, R.; Chen, J.; Wu, X.; Zhao, J. A correlation research of Ki67 index, CT features, and risk stratification in gastrointestinal stromal tumor. *Cancer Med* 2018, *7*, 4467-4474, doi:10.1002/cam4.1737.
44. Li, C.; Fu, W.; Huang, L.; Chen, Y.; Xiang, P.; Guan, J.; Sun, C. A CT-based nomogram for predicting the malignant potential of primary gastric gastrointestinal stromal tumors preoperatively. *Abdom Radiol (NY)* 2021, *46*, 3075-3085.
45. Liu, S.; Pan, X.; Liu, R.; Zheng, H.; Chen, L.; Guan, W.; Wang, H.; Sun, Y.; Tang, L.; Guan, Y.; et al. Texture analysis of CT images in predicting malignancy risk of gastrointestinal stromal tumours. *Clin Radiol* 2018, *73*, 266-274, doi:10.1016/j.crad.2017.09.003.
46. Mazzei, M.A.; Cioffi Squitieri, N.; Vindigni, C.; Guerrini, S.; Gentili, F.; Sadotti, G.; Mercuri, P.; Righi, L.; Lucii, G.; Mazzei, F.G.; et al. Gastrointestinal stromal tumors (GIST): a proposal of a "CT-based predictive model of Miettinen index" in predicting the risk of malignancy. *Abdom Radiol (NY)* 2020, *45*, 2989-2996, doi:10.1007/s00261-019-02209-7.

47. Miyake, K.K.; Nakamoto, Y.; Mikami, Y.; Tanaka, S.; Higashi, T.; Tadamura, E.; Saga, T.; Minami, S.; Togashi, K. The predictive value of preoperative (18)F-fluorodeoxyglucose PET for postoperative recurrence in patients with localized primary gastrointestinal stromal tumour. *Eur Radiol* 2016, *26*, 4664-4674, doi:10.1007/s00330-016-4242-5.
48. O'Neill, A.C.; Shinagare, A.B.; Kurra, V.; Tirumani, S.H.; Jagannathan, J.P.; Baheti, A.D.; Hornick, J.L.; George, S.; Ramaiya, N.H. Assessment of metastatic risk of gastric GIST based on treatment-naïve CT features. *Eur J Surg Oncol* 2016, *42*, 1222-1228, doi:10.1016/j.ejso.2016.03.032.
49. Pelandre, G.L.; Djahjah, M.C.; Gasparetto, E.L.; Nacif, M.S.; Marchiori, E.; De Mello, E.L.R. To mographic findings of gastric gastrointestinal stromal tumor and correlation with the mitotic index. *Arquivos de Gastroenterologia* 2013, *50*, 244-250, doi:http://dx.doi.org/10.1590/s0004-28032013000400002.
50. Palatresi, D.; Fedeli, F.; Danti, G.; Pasqualini, E.; Castiglione, F.; Messerini, L.; Massi, D.; Bettarini, S.; Tortoli, P.; Busoni, S.; et al. Correlation of CT radiomic features for GISTs with pathological classification and molecular subtypes: preliminary and monocentric experience. *Radiol Med* 2022, *127*, 117-128, doi:10.1007/s11547-021-01446-5.
51. Peng, G.; Huang, B.; Yang, X.; Pang, M.; Li, N. Preoperative CT feature of incomplete overlying enhancing mucosa as a high-risk predictor in gastrointestinal stromal tumors of the stomach. *Eur Radiol* 2021, *31*, 3276-3285.
52. Pinaikul, S.; Woodtichartpreecha, P.; Kanngurn, S.; Leelakiatpaiboon, S. 1189 Gastrointestinal stromal tumor (GIST): computed tomographic features and correlation of CT findings with histologic grade. *J Med Assoc Thai* 2014, *97*, 1189-1198.
53. Tang, B.; Feng, Q.X.; Liu, X.S. Comparison of Computed Tomography Features of Gastric and Small Bowel Gastrointestinal Stromal Tumors With Different Risk Grades. *J Comput Assist Tomogr* 2022, *46*, 175-182, doi:10.1097/rct.0000000000001262.
54. Tateishi, U.; Hasegawa, T.; Satake, M.; Moriyama, N. Gastrointestinal stromal tumor. Correlation of computed tomography findings with tumor grade and mortality. *J Comput Assist Tomogr* 2003, *27*, 792-798, doi:10.1097/00004728-200309000-00018.
55. Tokumoto, N.; Tanabe, K.; Misumi, T.; Fujikuni, N.; Suzuki, T.; Ohdan, H. The usefulness of preoperative 18FDG positron-emission tomography and computed tomography for predicting the malignant potential of gastrointestinal stromal tumors. *Dig Surg* 2014, *31*, 79-86, doi:10.1159/000357149.
56. Ulusan, S.; Koc, Z.; Kayaselcuk, F. Gastrointestinal stromal tumours: CT findings. *Br J Radiol* 2008, *81*, 618-623, doi:10.1259/bjr/90134736.
57. Verde, F.; Hruban, R.H.; Fishman, E.K. Small Bowel Gastrointestinal Stromal Tumors: Multidetector Computed Tomography Enhancement Pattern and Risk of Progression. *J Comput Assist Tomogr* 2017, *41*, 407-411, doi:10.1097/rct.0000000000000526.
58. Wei, S.C.; Xu, L.; Li, W.H.; Li, Y.; Guo, S.F.; Sun, X.R.; Li, W.W. Risk stratification in GIST: shape quantification with CT is a predictive factor. *Eur Radiol* 2020, *30*, 1856-1865, doi:10.1007/s00330-019-06561-6.
59. Xu, D.; Si, G.Y.; He, Q.Z. Correlation analysis of multi-slice computed tomography (MSCT) findings, clinicopathological factors, and prognosis of gastric gastrointestinal stromal tumors. *Transl Cancer Res* 2020, *9*, 1787-1794, doi:10.21037/tcr.2020.02.26.
60. Yang, T.H.; Hwang, J.I.; Yang, M.S.; Hung, S.W.; Chan, S.W.; Wang, J.; Tyan, Y.S. Gastrointestinal stromal tumors: computed tomographic features and prediction of malignant risk from computed tomographic imaging. *J Chin Med Assoc* 2007, *70*, 367-373, doi:10.1016/s1726-4901(08)70022-4.
61. Yang, C.W.; Liu, X.J.; Zhao, L.; Che, F.; Yin, Y.; Chen, H.J.; Zhang, B.; Wu, M.; Song, B. Preoperative prediction of gastrointestinal stromal tumors with high Ki-67 proliferation index based on CT features. *Ann Transl Med* 2021, *9*, 1556, doi:10.21037/atm-21-4669.
62. Yoshikawa, K.; Shimada, M.; Kurita, N.; Sato, H.; Iwata, T.; Morimoto, S.; Miyatani, T.; Kashihara, H.; Takasu, C.; Matsumoto, N. Efficacy of PET-CT for predicting the malignant potential of gastrointestinal stromal tumors. *Surg Today* 2013, *43*, 1162-1167, doi:10.1007/s00595-012-0411-6.

63. Zhou, C.; Duan, X.; Zhang, X.; Hu, H.; Wang, D.; Shen, J. Predictive features of CT for risk stratifications in patients with primary gastrointestinal stromal tumour. *Eur Radiol* 2016, *26*, 3086-3093, doi:10.1007/s00330-015-4172-7.
64. Zhu, M.P.; Ding, Q.L.; Xu, J.X.; Jiang, C.Y.; Wang, J.; Wang, C.; Yu, R.S. Building contrast-enhanced CT-based models for preoperatively predicting malignant potential and Ki67 expression of small intestine gastrointestinal stromal tumors (GISTs). *Abdom Radiol (NY)* 2021, doi:10.1007/s00261-021-03040-9.
65. Otomi, Y.; Otsuka, H.; Morita, N.; Terazawa, K.; Furutani, K.; Harada, M.; Nishitani, H. Relationship between FDG uptake and the pathological risk category in gastrointestinal stromal tumors. *J Med Invest* 2010, *57*, 270-274, doi:10.2152/jmi.57.270.
66. Park, J.-W.; Cho, C.-H.; Jeong, D.-S.; Chae, H.-D. Role of F-fluoro-2-deoxyglucose Positron Emission Tomography in Gastric GIST: Predicting Malignant Potential Pre-operatively. *J Gastric Cancer* 2011, *11*, 173-179, doi:10.5230/jgc.2011.11.3.173.
67. Yin, Y.Q.; Liu, C.J.; Zhang, B.; Wen, Y.; Yin, Y. Association between CT imaging features and KIT mutations in small intestinal gastrointestinal stromal tumors. *Sci Rep* 2019, *9*, 7257, doi:10.1038/s41598-019-43659-9.
68. Wang, J.K. Predictive value and modeling analysis of MSCT signs in gastrointestinal stromal tumors (GISTs) to pathological risk degree. *Eur Rev Med Pharmacol Sci* 2017, *21*, 999-1005.
69. Ao, W.; Cheng, G.; Lin, B.; Yang, R.; Liu, X.; Zhou, C.; Wang, W.; Fang, Z.; Tian, F.; Yang, G.; et al. A novel CT-based radiomic nomogram for predicting the recurrence and metastasis of gastric stromal tumors. *American Journal of Cancer Research* 2021, *11*, 3123-3134.
70. Chen, T.; Liu, S.; Li, Y.; Feng, X.; Xiong, W.; Zhao, X.; Yang, Y.; Zhang, C.; Hu, Y.; Chen, H.; et al. Developed and validated a prognostic nomogram for recurrence-free survival after complete surgical resection of local primary gastrointestinal stromal tumors based on deep learning. *EBioMedicine* 2019, *39*, 272-279, doi:10.1016/j.ebiom.2018.12.028.
71. Chen, T.; Ning, Z.; Xu, L.; Feng, X.; Han, S.; Roth, H.R.; Xiong, W.; Zhao, X.; Hu, Y.; Liu, H.; et al. Radiomics nomogram for predicting the malignant potential of gastrointestinal stromal tumours preoperatively. *Eur Radiol* 2019, *29*, 1074-1082, doi:10.1007/s00330-018-5629-2.
72. Chen, Z.H.; Xu, L.Y.; Zhang, C.M.; Huang, C.C.; Wang, M.H.; Feng, Z.; Xiong, Y. CT Radiomics Model for Discriminating the Risk Stratification of Gastrointestinal Stromal Tumors: A Multi-Class Classification and Multi-Center Study. *FRONTIERS IN ONCOLOGY* 2021, *11*.
73. Chu, H.; Pang, P.; He, J.; Zhang, D.; Zhang, M.; Qiu, Y.; Li, X.; Lei, P.; Fan, B.; Xu, R. Value of radiomics model based on enhanced computed tomography in risk grade prediction of gastrointestinal stromal tumors. *Sci Rep* 2021, *11*, 12009, doi:10.1038/s41598-021-91508-5.
74. Kang, B.; Yuan, X.; Wang, H.; Qin, S.; Song, X.; Yu, X.; Zhang, S.; Sun, C.; Zhou, Q.; Wei, Y.; et al. Preoperative CT-Based Deep Learning Model for Predicting Risk Stratification in Patients With Gastrointestinal Stromal Tumors. *Front Oncol* 2021, *11*, 750875, doi:10.3389/fonc.2021.750875.
75. Liu, B.; Liu, H.; Zhang, L.; Song, Y.; Yang, S.; Zheng, Z.; Zhao, J.; Hou, F.; Zhang, J. Value of contrast-enhanced CT based radiomic machine learning algorithm in differentiating gastrointestinal stromal tumors with KIT exon 11 mutation: a two-center study. *Diagn Interv Radiol* 2022, *28*, 29-38, doi:10.5152/dir.2021.21600.
76. Liu, X.; Yin, Y.; Wang, X.; Yang, C.; Wan, S.; Yin, X.; Wu, T.; Chen, H.; Xu, Z.; Li, X.; et al. Gastrointestinal stromal tumors: associations between contrast-enhanced CT images and KIT exon 11 gene mutation. *Ann Transl Med* 2021, *9*, 1496, doi:10.21037/atm-21-3811.
77. Feng, Q.; Tang, B.; Zhang, Y.; Liu, X. Prediction of the Ki-67 expression level and prognosis of gastrointestinal stromal tumors based on CT radiomics nomogram. *Int J Comput Assist Radiol Surg* 2022, doi:10.1007/s11548-022-02575-6.
78. Shao, M.; Niu, Z.; He, L.; Fang, Z.; He, J.; Xie, Z.; Cheng, G.; Wang, J. Building Radiomics Models Based on Triple-Phase CT Images Combining Clinical Features for Discriminating the Risk Rating in Gastrointestinal Stromal Tumors. *Front Oncol* 2021, *11*, 737302, doi:10.3389/fonc.2021.737302.
79. Ren, C.; Wang, S.; Zhang, S. Development and validation of a nomogram based on CT images and 3D texture analysis for preoperative prediction of the malignant potential in gastrointestinal stromal tumors. *Cancer Imaging* 2020, *20*, 5, doi:10.1186/s40644-019-0284-7.

80. Ren, C.; Wang, S.; Zhang, S.; Jiang, Z. Value of CT-Based Texture Analysis in Preoperative Prediction of the Grade of Gastrointestinal Stromal Tumors Compared to Conventional CT Imaging. *Iranian Journal of Radiology* 2019, *16*, e85703, doi:http://dx.doi.org/10.5812/iranjradiol.85703.
81. Starmans, M.P.A.; Timbergen, M.J.M.; Vos, M.; Renckens, M.; Grünhagen, D.J.; van Leenders, G.; Dwarkasing, R.S.; Willemssen, F.; Niessen, W.J.; Verhoef, C.; et al. Differential Diagnosis and Molecular Stratification of Gastrointestinal Stromal Tumors on CT Images Using a Radiomics Approach. *J Digit Imaging* 2022, *35*, 127-136, doi:10.1007/s10278-022-00590-2.
82. Wang, C.; Li, H.; Jiaerken, Y.; Huang, P.; Sun, L.; Dong, F.; Huang, Y.; Dong, D.; Tian, J.; Zhang, M. Building CT Radiomics-Based Models for Preoperatively Predicting Malignant Potential and Mitotic Count of Gastrointestinal Stromal Tumors. *Transl Oncol* 2019, *12*, 1229-1236, doi:10.1016/j.tranon.2019.06.005.
83. Wang, M.; Feng, Z.; Zhou, L.; Zhang, L.; Hao, X.; Zhai, J. Computed-Tomography-Based Radiomics Model for Predicting the Malignant Potential of Gastrointestinal Stromal Tumors Preoperatively: A Multi-Classifier and Multicenter Study. *Front Oncol* 2021, *11*, 582847, doi:10.3389/fonc.2021.582847.
84. Xu, F.; Ma, X.; Wang, Y.; Tian, Y.; Tang, W.; Wang, M.; Wei, R.; Zhao, X. CT texture analysis can be a potential tool to differentiate gastrointestinal stromal tumors without KIT exon 11 mutation. *Eur J Radiol* 2018, *107*, 90-97.
85. Xu, J.; Zhou, J.; Wang, X.; Fan, S.; Huang, X.; Xie, X.; Yu, R. A multi-class scoring system based on CT features for preoperative prediction in gastric gastrointestinal stromal tumors. *Am J Cancer Res* 2020, *10*, 3867-3881.
86. Zhang, L.; Kang, L.; Li, G.; Zhang, X.; Ren, J.; Shi, Z.; Li, J.; Yu, S. Computed tomography-based radiomics model for discriminating the risk stratification of gastrointestinal stromal tumors. *Radiol Med* 2020, *125*, 465-473, doi:10.1007/s11547-020-01138-6.
87. Zhang, Q.W.; Zhou, X.X.; Zhang, R.Y.; Chen, S.L.; Liu, Q.; Wang, J.; Zhang, Y.; Lin, J.; Xu, J.R.; Gao, Y.J.; et al. Comparison of malignancy-prediction efficiency between contrast and non-contract CT-based radiomics features in gastrointestinal stromal tumors: A multicenter study. *Clin Transl Med* 2020, *10*, e291.
88. Zhang, Q.W.; Gao, Y.J.; Zhang, R.Y.; Zhou, X.X.; Chen, S.L.; Zhang, Y.; Liu, Q.; Xu, J.R.; Ge, Z.Z. Personalized CT-based radiomics nomogram preoperative predicting Ki-67 expression in gastrointestinal stromal tumors: a multicenter development and validation cohort. *Clin Transl Med* 2020, *9*, 12, doi:10.1186/s40169-020-0263-4.
89. Zhao, Y.; Feng, M.; Wang, M.; Zhang, L.; Li, M.; Huang, C. CT Radiomics for the Preoperative Prediction of Ki67 Index in Gastrointestinal Stromal Tumors: A Multi-Center Study. *Front Oncol* 2021, *11*, 689136.
90. Zheng, J.; Xia, Y.; Xu, A.; Weng, X.; Wang, X.; Jiang, H.; Li, Q.; Li, F. Combined model based on enhanced CT texture features in liver metastasis prediction of high-risk gastrointestinal stromal tumors. *Abdom Radiol (NY)* 2022, *47*, 85-93, doi:10.1007/s00261-021-03321-3.
91. Antoch, G.; Kanja, J.; Bauer, S.; Kuehl, H.; Renzing-Koehler, K.; Schuette, J.; Bockisch, A.; Debatin, J.F.; Freudenberg, L.S. Comparison of PET, CT, and dual-modality PET/CT imaging for monitoring of imatinib (STI571) therapy in patients with gastrointestinal stromal tumors. *J Nucl Med* 2004, *45*, 357-365.
92. Beheshti, M.; Li, S.R.; Vali, R.; Schima, W.; Dudczak, R.; Langsteger, W. The Potential Value of F-18 FDG PET in Comparison to CT in Early Prediction of Response to Imatinib (STI571) Therapy in Patients with Gastrointestinal Stromal Tumors. *IRANIAN JOURNAL OF NUCLEAR MEDICINE* 2007, *15*, 34-42.
93. Chacón, M.; Eleta, M.; Espindola, A.R.; Roca, E.; Méndez, G.; Rojo, S.; Pupareli, C. Assessment of early response to imatinib 800 mg after 400 mg progression by ^{18}F-fluorodeoxyglucose PET in patients with metastatic gastrointestinal stromal tumors. *Future Oncol* 2015, *11*, 953-964, doi:10.2217/fon.14.292.
94. Choi, H.; Charnsangavej, C.; de Castro Faria, S.; Tamm, E.P.; Benjamin, R.S.; Johnson, M.M.; Macapinlac, H.A.; Podoloff, D.A. CT evaluation of the response of gastrointestinal stromal tumors

after imatinib mesylate treatment: a quantitative analysis correlated with FDG PET findings. *AJR Am J Roentgenol* 2004, *183*, 1619-1628, doi:10.2214/ajr.183.6.01831619.
95. Choi, H.; Charnsangavej, C.; Faria, S.C.; Macapinlac, H.A.; Burgess, M.A.; Patel, S.R.; Chen, L.L.; Podoloff, D.A.; Benjamin, R.S. Correlation of computed tomography and positron emission tomography in patients with metastatic gastrointestinal stromal tumor treated at a single institution with imatinib mesylate: proposal of new computed tomography response criteria. *J Clin Oncol* 2007, *25*, 1753-1759, doi:10.1200/jco.2006.07.3049.
96. Dudeck, O.; Zeile, M.; Reichardt, P.; Pink, D. Comparison of RECIST and Choi criteria for computed tomographic response evaluation in patients with advanced gastrointestinal stromal tumor treated with sunitinib. *Ann Oncol* 2011, *22*, 1828-1833, doi:10.1093/annonc/mdq696.
97. Gayed, I.; Vu, T.; Iyer, R.; Johnson, M.; Macapinlac, H.; Swanston, N.; Podoloff, D. The role of 18F-FDG PET in staging and early prediction of response to therapy of recurrent gastrointestinal stromal tumors. *J Nucl Med* 2004, *45*, 17-21.
98. Goerres, G.W.; Stupp, R.; Barghouth, G.; Hany, T.F.; Pestalozzi, B.; Dizendorf, E.; Schnyder, P.; Luthi, F.; von Schulthess, G.K.; Leyvraz, S. The value of PET, CT and in-line PET/CT in patients with gastrointestinal stromal tumours: long-term outcome of treatment with imatinib mesylate. *Eur J Nucl Med Mol Imaging* 2005, *32*, 153-162, doi:10.1007/s00259-004-1633-7.
99. Holdsworth, C.H.; Badawi, R.D.; Manola, J.B.; Kijewski, M.F.; Israel, D.A.; Demetri, G.D.; Van den Abbeele, A.D. CT and PET: early prognostic indicators of response to imatinib mesylate in patients with gastrointestinal stromal tumor. *AJR Am J Roentgenol* 2007, *189*, W324-330, doi:10.2214/ajr.07.2496.
100. Jager, P.L.; Gietema, J.A.; van der Graaf, W.T. Imatinib mesylate for the treatment of gastrointestinal stromal tumours: best monitored with FDG PET. *Nucl Med Commun* 2004, *25*, 433-438, doi:10.1097/00006231-200405000-00002.
101. Phongkitkarun, S.; Phaisanphrukkun, C.; Jatchavala, J.; Sirachainan, E. Assessment of gastrointestinal stromal tumors with computed tomography following treatment with imatinib mesylate. *World J Gastroenterol* 2008, *14*, 892-898, doi:10.3748/wjg.14.892.
102. Prior, J.O.; Montemurro, M.; Orcurto, M.V.; Michielin, O.; Luthi, F.; Benhattar, J.; Guillou, L.; Elsig, V.; Stupp, R.; Delaloye, A.B.; et al. Early prediction of response to sunitinib after imatinib failure by 18F-fluorodeoxyglucose positron emission tomography in patients with gastrointestinal stromal tumor. *J Clin Oncol* 2009, *27*, 439-445, doi:10.1200/jco.2008.17.2742.
103. Ryu, M.H.; Lee, J.L.; Chang, H.M.; Kim, T.W.; Kang, H.J.; Sohn, H.J.; Lee, J.S.; Kang, Y.K. Patterns of progression in gastrointestinal stromal tumor treated with imatinib mesylate. *Jpn J Clin Oncol* 2006, *36*, 17-24, doi:10.1093/jjco/hyi212.
104. Schindler, E.; Amantea, M.A.; Karlsson, M.O.; Friberg, L.E. PK-PD modeling of individual lesion FDG-PET response to predict overall survival in patients with sunitinib-treated gastrointestinal stromal tumor. *CPT Pharmacometrics Syst Pharmacol* 2016, *5*, 173-181, doi:10.1002/psp4.12057.
105. Schramm, N.; Englhart, E.; Schlemmer, M.; Hittinger, M.; Übleis, C.; Becker, C.R.; Reiser, M.F.; Berger, F. Tumor response and clinical outcome in metastatic gastrointestinal stromal tumors under sunitinib therapy: comparison of RECIST, Choi and volumetric criteria. *Eur J Radiol* 2013, *82*, 951-958, doi:10.1016/j.ejrad.2013.02.034.
106. Shinagare, A.B.; Barysauskas, C.M.; Braschi-Amirfarzan, M.; O'Neill, A.C.; Catalano, P.J.; George, S.; Ramaiya, N.H. Comparison of performance of various tumor response criteria in assessment of sunitinib activity in advanced gastrointestinal stromal tumors. *Clin Imaging* 2016, *40*, 880-884, doi:10.1016/j.clinimag.2016.04.007.
107. Van den Abbeele, A.D.; Gatsonis, C.; de Vries, D.J.; Melenevsky, Y.; Szot-Barnes, A.; Yap, J.T.; Godwin, A.K.; Rink, L.; Huang, M.; Blevins, M.; et al. ACRIN 6665/RTOG 0132 phase II trial of neoadjuvant imatinib mesylate for operable malignant gastrointestinal stromal tumor: monitoring with 18F-FDG PET and correlation with genotype and GLUT4 expression. *J Nucl Med* 2012, *53*, 567-574, doi:10.2967/jnumed.111.094425.
108. Stroobants, S.; Goeminne, J.; Seegers, M.; Dimitrijevic, S.; Dupont, P.; Nuyts, J.; Martens, M.; van den Borne, B.; Cole, P.; Sciot, R.; et al. 18FDG-Positron emission tomography for the early

prediction of response in advanced soft tissue sarcoma treated with imatinib mesylate (Glivec). *Eur J Cancer* 2003, *39*, 2012-2020, doi:10.1016/s0959-8049(03)00073-x.
109. Schiavon, G.; Ruggiero, A.; Bekers, D.J.; Barry, P.A.; Sleijfer, S.; Kloth, J.; Krestin, G.P.; Schöffski, P.; Verweij, J.; Mathijssen, R.H. The effect of baseline morphology and its change during treatment on the accuracy of Response Evaluation Criteria in Solid Tumours in assessment of liver metastases. *Eur J Cancer* 2014, *50*, 972-980, doi:10.1016/j.ejca.2014.01.004.
110. Schiavon, G.; Ruggiero, A.; Schöffski, P.; van der Holt, B.; Bekers, D.J.; Eechoute, K.; Vandecaveye, V.; Krestin, G.P.; Verweij, J.; Sleijfer, S.; et al. Tumor volume as an alternative response measurement for imatinib treated GIST patients. *PLoS One* 2012, *7*, e48372, doi:10.1371/journal.pone.0048372.
111. Farag, S.; Geus-Oei, L.F.; van der Graaf, W.T.; van Coevorden, F.; Grunhagen, D.; Reyners, A.K.L.; Boonstra, P.A.; Desar, I.; Gelderblom, H.; Steeghs, N. Early Evaluation of Response Using (18)F-FDG PET Influences Management in Gastrointestinal Stromal Tumor Patients Treated with Neoadjuvant Imatinib. *J Nucl Med* 2018, *59*, 194-196.
112. Farag, S.; NS, I.J.; Houdijk, M.P.M.; Reyners, A.K.L.; Arens, A.I.; Grünhagen, D.J.; Desar, I.M.E.; Gelderblom, H.; Steeghs, N.; de Geus-Oei, L.F. Early response evaluation using 18F-FDG-PET/CT does not influence management of patients with metastatic gastrointestinal stromal tumors (GIST) treated with palliative intent. *Nuklearmedizin* 2021, *60*, 411-416, doi:10.1055/a-1542-6211.
113. Goh, B.K.; Chow, P.K.; Chuah, K.L.; Yap, W.M.; Wong, W.K. Pathologic, radiologic and PET scan response of gastrointestinal stromal tumors after neoadjuvant treatment with imatinib mesylate. *Eur J Surg Oncol* 2006, *32*, 961-963, doi:10.1016/j.ejso.2006.06.004.
114. Arshad, J.; Ahmed, J.; Subhawong, T.; Trent, J.C. Progress in determining response to treatment in gastrointestinal stromal tumor. *EXPERT REVIEW OF ANTICANCER THERAPY* 2020, *20*, 279-288, doi:10.1080/14737140.2020.1745068.
115. Padhani, A.R.; Ollivier, L. The RECIST (Response Evaluation Criteria in Solid Tumors) criteria: implications for diagnostic radiologists. *Br J Radiol* 2001, *74*, 983-986, doi:10.1259/bjr.74.887.740983.
116. Faivre, S.; Demetri, G.; Sargent, W.; Raymond, E. Molecular basis for sunitinib efficacy and future clinical development. *Nat Rev Drug Discov* 2007, *6*, 734-745, doi:10.1038/nrd2380.
117. Jin, T.; Nakatani, H.; Taguchi, T.; Nakano, T.; Okabayashi, T.; Sugimoto, T.; Kobayashi, M.; Araki, K. STI571 (Glivec) suppresses the expression of vascular endothelial growth factor in the gastrointestinal stromal tumor cell line, GIST-T1. *World J Gastroenterol* 2006, *12*, 703-708, doi:10.3748/wjg.v12.i5.703.
118. Young, H.; Baum, R.; Cremerius, U.; Herholz, K.; Hoekstra, O.; Lammertsma, A.A.; Pruim, J.; Price, P. Measurement of clinical and subclinical tumour response using [18F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET Study Group. *Eur J Cancer* 1999, *35*, 1773-1782, doi:10.1016/s0959-8049(99)00229-4.

2

RECIST, Choi criteria or volumetry: early response assessment in gastrointestinal stromal tumour patients treated with neoadjuvant systemic treatment

RECIST, Choi criteria or volumetry: early response assessment in gastrointestinal stromal tumour patients treated with neoadjuvant systemic treatment

Ylva A. Weeda[1], Gijsbert M. Kalisvaart[1], Aart J. van der Molen[1], Floris H.P. van Velden[1], Daphne D.D. Rietbergen[1], Lioe-Fee de Geus-Oei[1,2,3], Willem Grootjans[1], Jos A. van der Hage[4]

[1]Department of Radiology, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands.
[2]Biomedical Photonic Imaging Group, University of Twente, 7522 NB Enschede, The Netherlands
[3]Department of Radiation Science & Technology, Technical University of Delft, 2629 JB Delft, The Netherlands
[4]Department of Surgical Oncology, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands.

Abstract: Neoadjuvant tyrosine kinase inhibitor (TKI) therapy is administered in gastrointestinal stromal tumour (GIST) patients to attain size reduction of the primary tumour and make complete resection feasible and less invasive. Response to therapy is typically determined through quantitative radiographic response criteria, involving tumour size and density measurements. However, it is still unclear whether these criteria can be used to determine the surgical benefit that can be achieved from neoadjuvant TKI treatment. This single-centre retrospective study aims to determine whether surgical benefit after neoadjuvant TKI treatment may be predicted based on quantitative radiological response criteria, i.e., RECIST 1.1, Choi criteria and volumetry. A total of 38 non-metastatic GIST patients were treated with neoadjuvant TKI treatment, followed by curative-intent surgery and monitored using contrast-enhanced computed tomography (CE-CT). Each tumour was manually delineated, and the longest transaxial diameter, volume and mean tumour density were automatically calculated. The surgical benefit was retrospectively determined by a surgical oncologist specialised in GIST management, blinded for radiological response criteria, and based on surgical and radiological reports and predefined criteria. Surgical benefit was defined as TKI therapy-induced tumour size reductions facilitating tissue and organ preservation or improved surgical oversight and planning. Patients were treated with neoadjuvant TKI therapy for a median interval of 284 days. Twenty-three out of 38 patients were scored to have a surgical benefit. When compared to the findings on surgical benefit, accuracy, sensitivity and specificity for RECIST 1.1 (76.3%, 93.3% and 65.2%, respectively), Choi (68.4%, 26.7%, and 95.7%, respectively) and volumetry (86.6%, 93.3%, and 82.6%, respectively) were calculated. Size-based criteria (RECIST 1.1 and volumetry) are most appropriate for neoadjuvant response assessment, as these reflect the obtained surgical benefit. Size measurements are less prone to scanner and imaging protocol variabilities when compared to density measurements, as presented in the Choi criteria. Future research should focus on standardisation of CE-CT acquisition protocols and the development of response prediction models that assist in making earlier image-guided treatment decisions.

1. Introduction

Gastrointestinal stromal tumours (GISTs) are rare mesenchymal neoplasms with a worldwide incidence of one or two per 100,000 [1,2]. Even though these tumours are radio- and chemotherapy resistant, systematic treatment involves the use of tyrosine kinase inhibitor (TKI) therapy including imatinib, sunitinib and regorafenib [3,4]. Neoadjuvant TKI treatment is administered in a selected group of patients, to attain primary tumour size reduction and increase probability of complete excision, while preserving surrounding tissue and organs as much as possible [5-7]. Clinical genotyping of the primary tumour through biopsy, is essential for making decisions regarding neoadjuvant therapy, since the sensitivity and resistance towards TKI treatment in GISTs is highly dependent on the mutational status [8].

Whether neoadjuvant treatment induces reduction in tumour size is, however, difficult to predict. Currently, the Response Evaluation Criteria in Solid Tumour (RECIST 1.1) are the gold standard for determining responder status in GIST patients, which measure changes in tumour size on two consecutive contrast-enhanced computed tomography (CE-CT) scans [9]. In some cases, TKI treatment can cause tumour destruction without inducing substantial tumour shrinkage and size-based criteria may therefore underestimate the therapeutic effect in GISTs [10,11]. In order to more accurately determine treatment response, changes in tumour density on portal contrast-enhanced CT imaging (Choi criteria) have been used to monitor the antiangiogenic effect of TKI treatment [12-14]. However, this technique is sensitive to variations in scanner and imaging protocols, in particular the timing of imaging after administration of iodinated contrast [15]. Furthermore, it remains questionable whether these quantitative radiographic response criteria translate to the added surgical benefits that can be induced by neoadjuvant TKI treatment.

Early and appropriate response prediction of treatment effectiveness and associated surgical benefit in neoadjuvant GIST treatment is crucial to prevent surgical delay and overtreatment, causing side effects and unnecessary costs in case of futile treatment. Moreover, tools for appropriate response assessment are necessary for identification of surrogate endpoints for imaging studies on therapy effect and treatment personalisation in this rare and heterogeneous disease. The purpose of this study is to determine whether surgical benefit in non-metastatic GIST can be correlated to current radiological criteria for response assessment.

2. Methods

2.1 Patients

This study retrospectively identified 57 patients with a confirmed non-metastatic primary GIST diagnosis, who were referred for neoadjuvant TKI treatment or follow-up in the Leiden University Medical Center from October 2003 until April 2022. Patients with non-metastatic GIST receiving neoadjuvant TKI treatment followed by curative-intended surgical resection, who were monitored using portal phase contrast-enhanced CT (CE-CT) imaging, were included. A total of 19 patients were eventually excluded due to missing imaging data (n = 4), concurrent treatment for a second malignancy (n = 2), refrainment from resection (n = 6) and absence of portal venous phase contrast (n = 6) or follow-up CT imaging (n = 1). Refrainment from surgery was attributed to personal preference of the patient, comorbidities or if the continuation of TKI treatment was preferred to prevent the need for a colostomy. Consequently, 38 GIST patients were included for final analysis (Fig. 1). The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee Leiden Den Haag Delft (METC LDD) (protocol code: B19.050, date of approval 14 January 2020). Patient consent was waived due to the retrospective nature of the study.

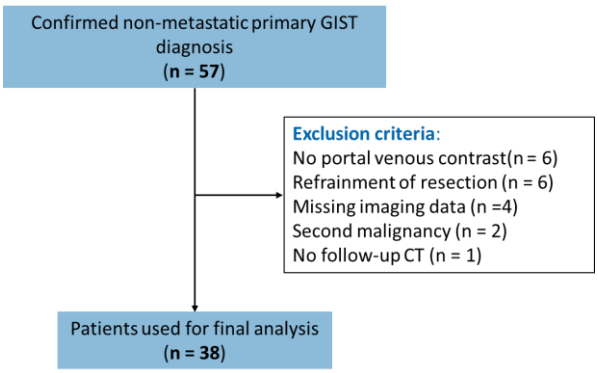

Fig. 1. Flow diagram visualising the selection of patients with predefined exclusion criteria. A total of 57 patients had a confirmed non-metastatic primary GIST diagnosis. After the selection procedure, 38 patients were used for the final analysis.

2.2 Surgical benefit

The patients were individually assessed to see whether the response to neoadjuvant TKI therapy did facilitate local surgical treatment. Surgical benefit was defined as TKI therapy-induced tumour size reduction facilitating tissue and organ preservation and/or facilitating the surgical procedure as a result of improved surgical oversight and planning. This was determined by a surgical oncologist with over ten years of surgical experience in the field of GISTs (JvdH). The surgeon was provided access to all imaging and clinicopathological records (reporting on pathology, radiology, surgery, multidisciplinary meetings, etc.) and was blinded for the calculated radiological response criteria in this study. Patients were considered to have obtained surgical benefit when TKI therapy-induced tumour size reduction ensured tissue and organ preservation and/or improved surgical oversight and planning.

2.3 Segmentation

Portal venous phase CE-CT imaging was collected for each individual patient, which was performed prior to neoadjuvant TKI administration and prior to surgery. The CE-CT images with the smallest slice thicknesses (0.5 - 5.0 mm) were used for the segmentation process. Tumour segmentation was performed by a technical medicine master student (YAW) using three-dimensional (3D) slicer (version 5.0.2) [16]. Lesions were manually delineated in the axial plane and automatically filled using slice-to-slice interpolation to create 3D tumour volumes. Segmentation was supervised by an expert abdominal radiologist with over 20 years of experience (AJvdM). In addition, a representative selection of the patient cases (five pre-therapy and five final response scans) was independently delineated by a medical PhD student (GMK) to assess inter-observer variability.

2.4 Radiological response assessment

Tumour size: The longest transaxial tumour diameter and total volume were automatically calculated for each tumour by assessing each individual slice. The percentage change in tumour diameter and subsequent volume between pre-therapy and final response scans were determined for each patient to evaluate therapeutic response. Patients were scored according to the RECIST 1.1 criteria, which introduces four types of response; complete response (clearance of all lesions), partial response (≥ 30% reduction in tumour size without progression), progressive disease (≥ 20% increase in tumour size) and stable disease (neither progressive disease nor partial response) [17]. A change in diameter translates to a larger change in spherical volume, resulting in volumetric thresholds of ≥ 66% reduction (complete response) and ≥ 73% increase (progressive disease). These thresholds for volumetric measurements can be deduced from the formula for the corresponding spherical volume of a diameter (Equation 1).

$$(1) \quad Vsphere = \frac{4}{3}\,\pi\, r^3$$

Tumour density: The tumour density of the entire tumour volume was also computed, quantified by the mean of all attenuation coefficients (Hounsfield Units). The percentage change in tumour density between two consecutive CE-CT scans was calculated. Patients were scored according to the Choi criteria, which takes into account tumour diameter as well as tumour density; complete response (clearance of all lesions), partial response (≥ 10% reduction in tumour size or ≥ 15% reduction in tumour density), progressive disease (≥ 10% increase in tumour size and < 15% reduction in tumour density) and stable disease (neither progressive disease nor partial response) [12].

All measurements were automatically computed using Pyradiomics software (version 3.0.1) in Python (version 3.7) [18]. Since the goal of neoadjuvant therapy is to make complete resection more feasible, good responders were considered to have complete or partial response, while patients with stable or progressive disease were classified as poor responders.

2.5 Statistical analysis

Differences between the surgical benefit and non-benefit group in baseline characteristics were assessed using Python (version 3.7). The Mann-Whitney U test and Fisher's exact test were performed for numerical data and categorical data, respectively. A p-value below 0.05 was considered statistically significant. The inter-observer agreement between the executed segmentations (on a subset of ten patients) was assessed using the intra-class correlation coefficient (ICC) and the Dice Similarity Coefficient (DSC). To evaluate the performance and efficacy of the current radiographic criteria, their findings on responder status were compared to the surgical benefit assessment with accuracy, sensitivity and specificity.

3. Results

3.1 Patient characteristics

The demographics and clinicopathological characteristics of the included patients are listed in Table 1. Patients initially received a daily imatinib mesylate dosage of 400 mg, except for one patient who received avapritinib (300 mg daily) at the hand of an underlying platelet-derived growth factor alpha (PDGFRA) exon 18 (D842V) mutation. Dosage was reduced in case of intolerable side-effects (n = 1) or reduced kidney function (n = 1) and escalated in patients with disease progression under treatment (n = 1) or low imatinib plasma levels (n = 2). The median interval between start of treatment and the last response scan was 212 days (range 29 – 699). Patients were referred for surgery in case of tumour progression or when maximal response was reached on two consecutive scans. A median interval of 50 days (range 3 - 167) was observed between the last response scan and surgery. TKI therapy was terminated just before surgery, which resulted in a median treatment duration of 284 days (range 52 – 702).

3.2 Surgical benefit

Neoadjuvant TKI treatment induced surgical benefit in a total of 23 patients. Significant changes in tumour size were observed in thirteen patients, which facilitated the surgical procedure, and/or enabled preservation of surrounding tissue and eliminated considerable risks (Fig. 2). In four of these patients, organ-preserving (e.g., spleen (n = 2), anal sphincter (n = 1), common bile duct and concurrent pancreatic duct (n = 1)) surgery was feasible. Nonetheless, surgical benefit was not only determined by significant reductions in tumour size. Also, more subtle and moderate changes in tumour size led to surgical benefit, due to provision of better surgical exposure and simpler surgical planning in ten patients. This was mainly attributed to the neutralisation of tumour adhesions to adjacent structures. Surgical procedures were for

instance facilitated by the need for smaller incisions or when a laparotomy could be converted to a laparoscopic procedure. In addition, in gastric GISTs, the surgeon opted for a wedge excision or a partial gastrectomy, based on the area of tumour attachment to the affected organ. Improved visualisation of tumour attachment was therefore imperative to surgical decision making as well (Fig. 3). There were no significant differences in baseline characteristics between the patients with or without surgical benefit (Table 1).

Patient characteristics				
	Total (n = 38)	Surgical benefit (n = 23)	No surgical benefit (n = 15)	p-value
Age (years)	62.0 (48.8 - 71.0)	61.0 (46.0 - 67.5)	68.0 (54.0 - 71.5)	0.49
Sex				0.06
Female	10 (26.3%)	9 (39.1%)	1 (6.7%)	
Male	28 (73.7%)	14 (65.2%)	14 (93.3%)	
Tumour location				1.00
Stomach	28 (73.7%)	17 (73. 9%)	11 (73.3%)	*Upper vs. lower GI tract*
Duodenum	5 (13.2%)	3 (13.0%)	2 (13.3%)	
Small intestine	2 (5.3%)	1 (4.4%)	1 (6.7%)	
Rectum	3 (7.9%)	2 (8.7%)	1 (6.7%)	
Largest primary tumour diameter (mm)	105.5 (76.0 - 192.5)	106.0 (81.0 - 185.0)	89.0 (61.5 - 189.0)	0.31
Mitotic index				0.63
≤ 5 per 5 mm^2	17 (44.7%)	13 (56.5%)	4 (26.7%)	*High vs. low mitotic index*
> 5 per 5 mm^2	6 (15.8%)	4 (17.4%)	2 (13.3%)	
Not reported	15 (39.5%)	6 (26.1%)	9 (60.0%)	
Risk stratifciation				0.62
Low	7 (18.4%)	4 (17.4%)	3 (20.0%)	*High vs. Intermediate and low risk*
Intermediate	8 (21.1%)	8 (34.8%)	0 (0.0%)	
High	8 (21.1%)	5 (21.7%)	3 (20.0%)	
Not reported	15 (39.5%)	6 (26.1%)	11 (73.3%)	
Mutational status				0.16
KIT exon 11	26 (68.4%)	18 (78.3%)	8 (53.3%)	*KIT exon 11 vs. other mutations*
KIT exon 9	2 (5.3%)	1 (4.4%)	1 (6.7%)	
PDGFRA exon 14	1 (2.6%)	0 (0.0%)	1 (6.7%)	
PDGFRA exon 18	7 (18.4%)	4 (17.4%)	3 (20.0%)	
Wildtype	2 (5.3%)	0 (0.0%)	2 (13.3%)	
Interval start therapy-last scan (days)	211.5 (108.0 - 324.0)	253 (199.5 - 328.5)	158 (88.5 - 237.4)	0.10
Interval last scan-surgery (days)	50 (36.0 - 74.0)	51 (39.0 - 71.0)	49 (31.0 - 78.0)	0.86
Treatment interval (days)	284.0 (160.5 - 372.8)	323 (262.0 - 378.5)	235 (141.0 - 299.5)	0.07

Table 1. Baseline characteristics of the included non-metastatic GIST patients, stratified by the assessed surgical benefit. Numerical data is presented as a median value, along with the concurrent interquartile range. Risk stratification was assessed using the Miettinen classification system involving mitotic index, tumour size and location as prognostic factors. Tumour size was in this case defined by the largest tumour dimension. P-values for numerical data were calculated through the Mann-Whitney and the Fisher's exact test was used for categorical data. (GI = gastrointestinal, PDGFRA = platelet-derived growth factor alpha, KIT = receptor tyrosine kinase). * Upper gastrointestinal tract (stomach and duodenum), lower gastrointestinal tract (small intestine and rectum). † Non-reported cases were omitted from statistical analysis.

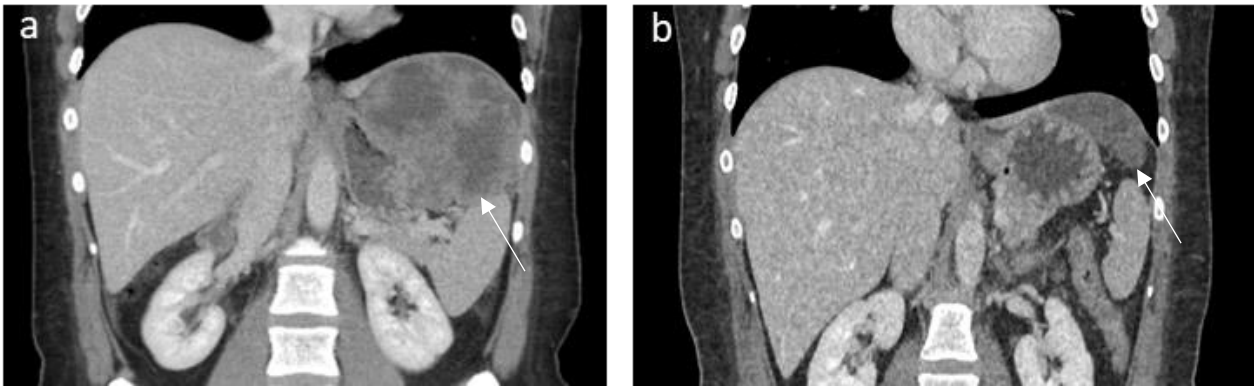

Fig. 2 a) Coronal portal venous phase CE-CT image of a 46-year-old female diagnosed with an intermediate-risk gastric GIST (arrow). b) A substantial reduction in tumour size (arrow) was observed after 11.8 months of neoadjuvant TKI treatment, which was considered surgically beneficial.

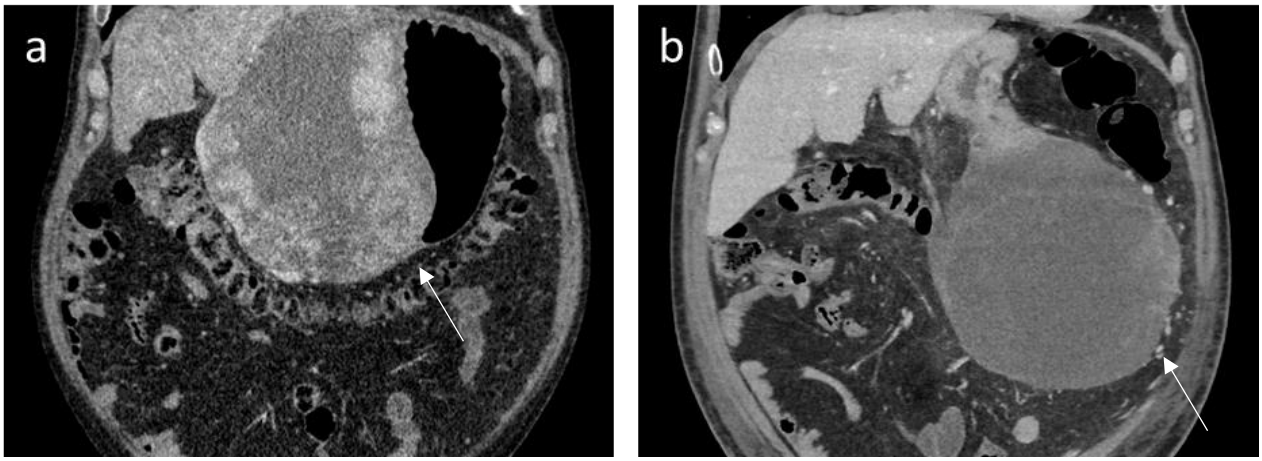

Fig. 3 Coronal portal venous phase CE-CT images of an 86-year-old male diagnosed with a gastric GIST (arrows). a) Before treatment induction, the origin of the tumour was difficult to assess. As a result, the surgeon opted for a partial gastrectomy. b) After 1.8 months of TKI treatment, the tumour dimensions remained relatively unchanged. However, the improved visualisation of the area of tumour attachment to the stomach is imperative for surgical planning and oversight. Based on this information, a partial gastrectomy could be converted to a wedge excision.

3.3 Image acquisition

Many diagnostic CE-CT images were obtained from other hospitals and consequently heterogeneity in CE-CT acquisition and reconstruction protocols was introduced. In pre-therapy imaging, the slice thickness ranged from 0.6 to 5.0 mm, from 100 to 140 peak kilovoltage (kVp) and pixel spacing from 0.62 to 0.94 mm. Besides, four scanner manufacturers were observed including Siemens Healthineers, GE Healthcare, Canon Medical Systems (previously known as Toshiba Medical Systems) and Philips Healthcare. Slice thickness ranged from 0.5 to 1.0 mm kVp ranged from 120 to 140 and pixel spacing from 0.69 and 0.98 mm. In the final preoperative response scans, two scanner manufacturers were observed, namely Toshiba Medical Systems and GE Healthcare.

3.4 Radiological response assessment

The ICC between the segmentations made by two observers for tumour size, volume and density were all above 0.99. The mean DSC score between the segmentations made by the two observers was 0.90 (± 0.08 SD). Given these outcomes, the quantitative features extracted from the CE-CT imaging were considered to have good reliability across the segmentations.

The Choi criteria found 33 partial responses (PR), four patients with stable disease and one with progressive disease (PD). The performance scores were lower, with an accuracy, sensitivity and specificity of 68.4%, 26.7% and 95.7%. Using the RECIST 1.1 criteria, 16 patients had a PR, 21 had stable disease and one patient had PD. When compared to the subjective findings of the surgical benefit assessment, outcome measures were 76.3%, 93.3% and 65.2% for accuracy, sensitivity and specificity, respectively. According to the volumetric criteria, 18 patients had a PR and 20 were considered to have stable disease. This resulted in the highest performance scores of 86.6% accuracy, 93.3% sensitivity and 82.6% specificity. Percentage change in diameter and density are plotted for each patient and categorized for surgical benefit of neoadjuvant TKI in Fig. 4. In this study, RECIST 1.1 classified three patients as stable disease, while the percentage change in tumour volume was above 66.0% and surgical benefit was observed in these patients (Fig 5 and 6).

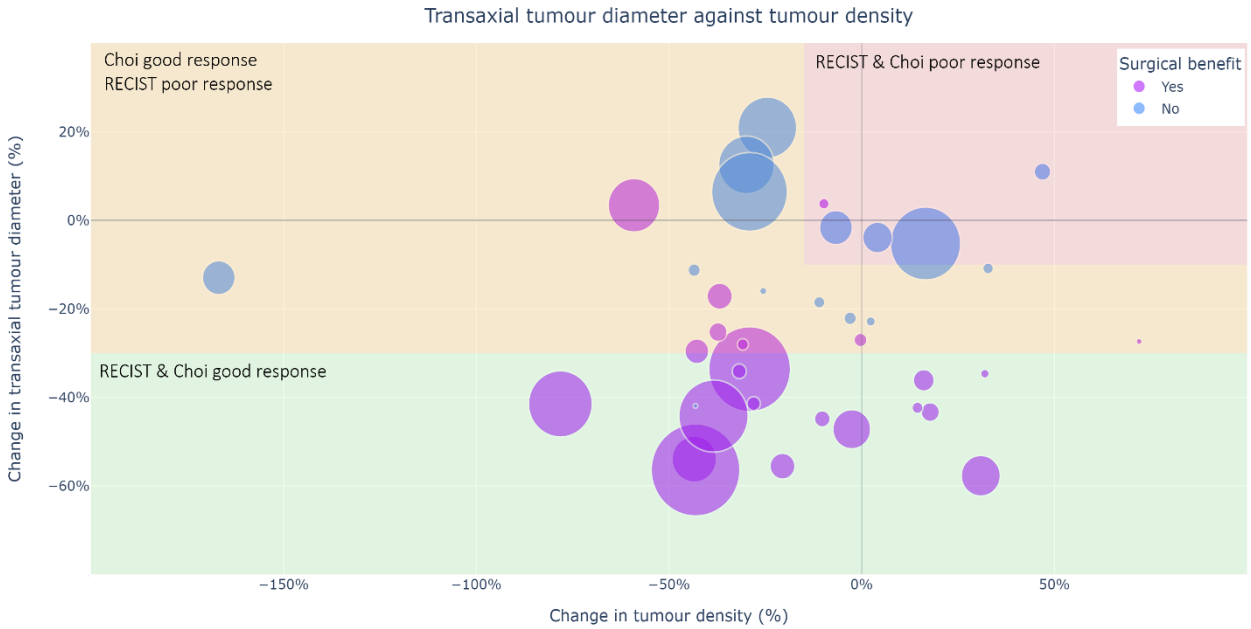

Fig. 4 Bubble plot visualising the percentage change in transaxial diameter against the percentage change in tumour density. Each bubble represents an individual non-metastatic GIST patient treated with neoadjuvant TKI treatment. The bubble colour resembles the assessment on surgical benefit (magenta = surgical benefit, blue = no surgical benefit,), while the bubble size represents the scaled initial tumour volume at diagnosis. The grid has been divided into three parts by using the thresholds presented in current radiographic response criteria (red = agreement on bad response, green = agreement on good response and orange = discrepancies on responder status).

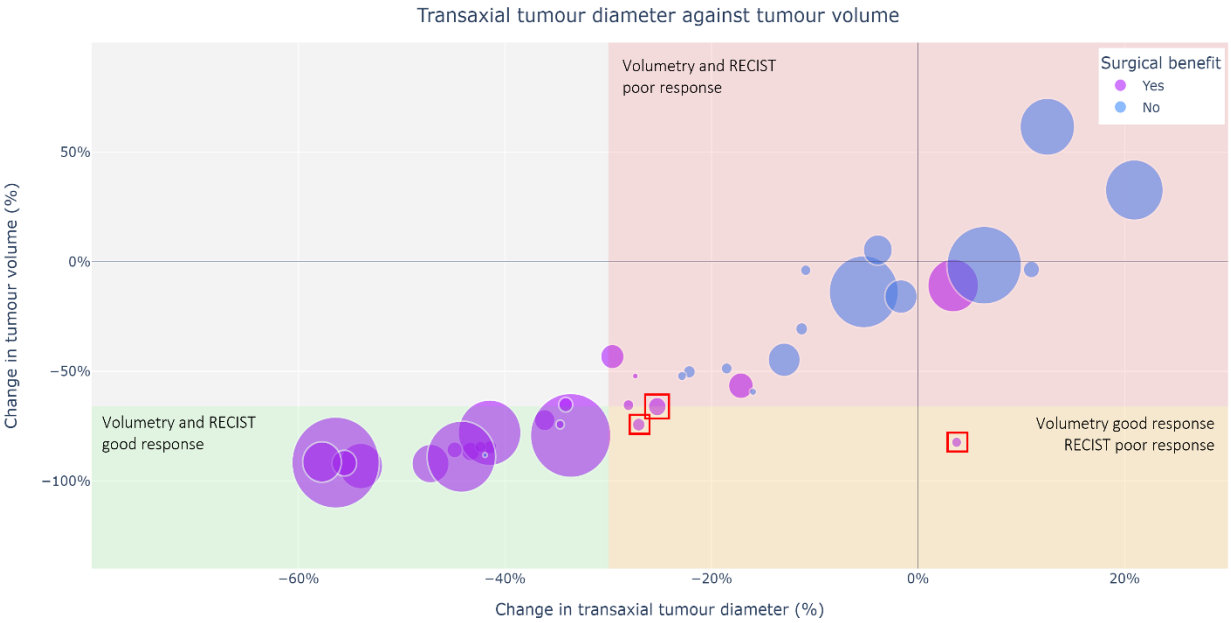

Fig. 5 Bubble plot visualising the percentage change in transaxial diameter against the percentage change in tumour volume. Each bubble represents an individual non-metastatic GIST patient treated with neoadjuvant TKI treatment. The bubble colour resembles the assessment on surgical benefit (magenta = surgical benefit, blue = no surgical benefit), while the bubble size represents the scaled initial tumour volume at diagnosis. The grid has been divided into three parts by using the thresholds presented in current radiographic response criteria (red = agreement on bad response, green = agreement on good response and orange = discrepancies on responder status). In this study, three patients were classified as non-responders by the RECIST 1.1 criteria, while the percentage change in tumour volume was above 66.0% and surgical benefit was observed in these patients (red boxes).

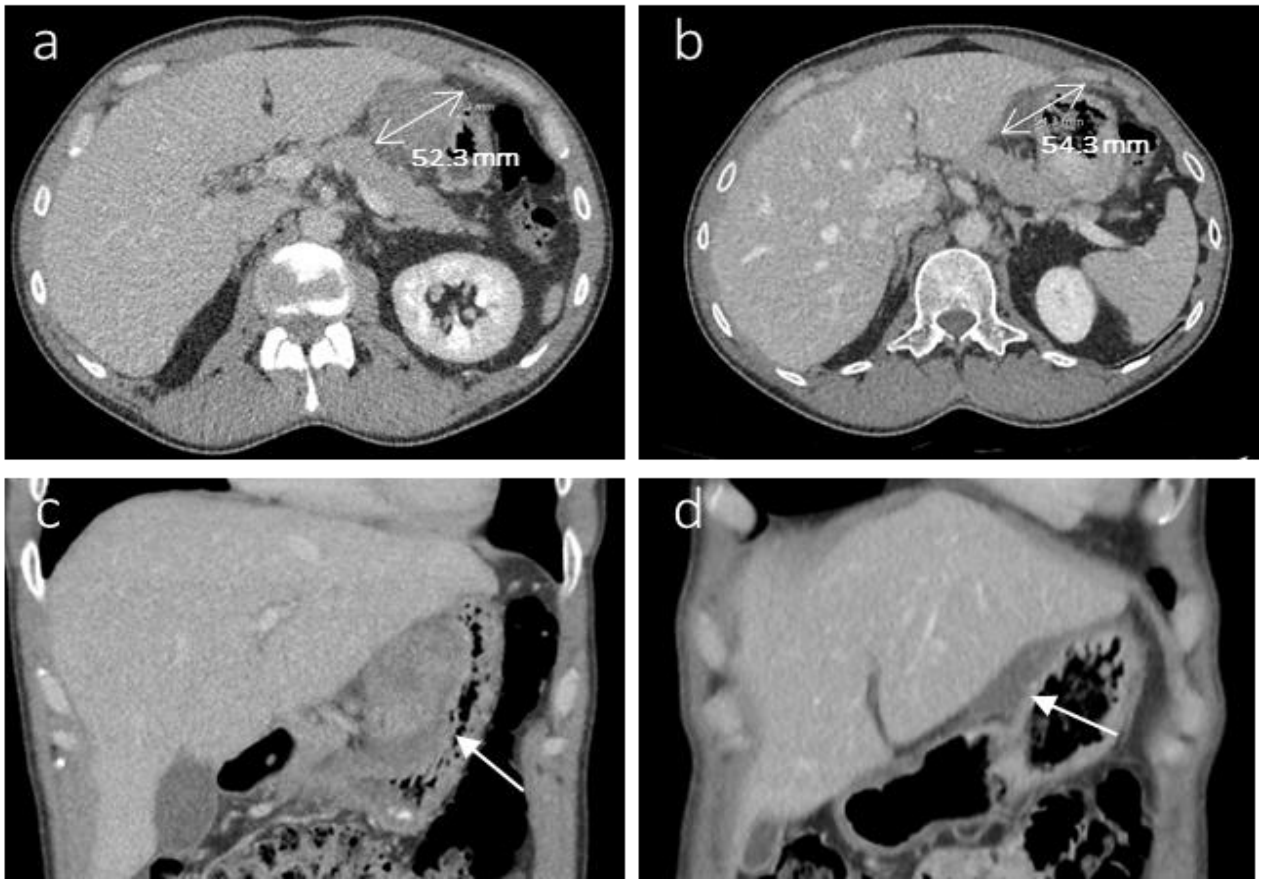

Fig. 6 a) Axial portal venous phase CE-CT image of a 45-year-old male diagnosed with a gastric GIST. Before treatment induction, the transaxial tumour diameter measured 52.3 mm. b) After 6.8 months of TKI treatment, the transaxial RECIST tumour diameter equalled 54.3 mm, suggesting minimal to no response to systemic treatment. However, when looking at the coronal portal venous CT image, a significant decrease in tumour volume (arrows) can be observed between the c) pre-therapy and d) final preoperative response scan. The percentage change in tumour volume equalled 82.4%. RECIST measurements underestimated the therapeutic effect of TKI treatment in this patient.

4. Discussion

In this study, the efficacy of current radiographic response criteria (i.e., RECIST 1.1, Choi and volumetry) in predicting response to neoadjuvant TKI treatment was evaluated by comparing quantitative findings on radiographic response status with the obtained surgical benefit. The results indicate that size-based criteria (RECIST 1.1 and volumetry) are clinically relevant for determining surgical benefit, rather than using changes in tumour density, as presented in the Choi criteria.

The diagnostic portal phase CE-CT images have been collected from multiple hospitals over a longer period of time. Non-uniform acquisition protocols and variable acquisition parameters across scanners induce variations in attenuation coefficients on CE-CT imaging [19-21]. The Choi criteria tries to reflect changes in biological tumour behaviour by measuring the percentage change in these attenuation coefficients. The induced inter- and intra-scanner variability makes side to side comparison on tumour density a difficult task with unreliable outcomes. In clinical practice, multi-centric data collection is, however, inevitable. The observed heterogeneity in acquisition and reconstruction protocols, and concurrent low predictive power in determining surgical benefit, make Choi criteria less practical and precise for clinical neoadjuvant response assessment in non-metastatic GIST patients.

RECIST 1.1 measurements accurately reflected surgical benefit, which was in line with results from Chen et al. The authors found a correlation between response determined through RECIST 1.1 measurements and a decreased scope of surgery in high risk soft tissue sarcomas [22]. Nonetheless, we hypothesise that morphological changes in GIST dimensions on consecutive CE-CT scans can be more accurately quantified by means of volumetry, compared to one-dimensional RECIST measurement. Using one-dimensional measurements, one presumes a constant size reduction along three orthogonal axes over the course of treatment. Schiavon et al. also showed that volumetric criteria classified a higher number of patients as partial responders compared to RECIST 1.1, which is in accordance with our findings [23]. However, before applying volumetric response assessment in clinical practice, a (semi-) automatic segmentation algorithm is preferred to avoid additional labour for radiologists and improve repeatability.

Different from many other tumours, there is no pathological gold standard for determining response in GIST patients. TKI treatment only sporadically induces complete pathological response and there is a lack of evidence for a possible correlation between the percentage of remaining viable tumour cells and overall survival [24]. As presented in the European Society for Medical Oncology (ESMO) guidelines, radiographic criteria are currently the only standards for monitoring tumour response in GISTs [8]. Despite these radiographic criteria, it remains challenging for physicians to determine surgical benefit preoperatively and decisions are mainly based on surgical experience. For less experienced physicians and surgeons in training, quantitative measures, such as tumour volume, will be helpful for assessing surgical benefit. However, more objective findings, such as better visualisation of the area of tumour attachment should be considered as well. We observed five patients that were considered to have obtained surgical benefit, while they were classified as non-responder by RECIST 1.1 and volumetric criteria. In three of these patients, tumour attachment played a crucial role in surgical decision making.

This study has some limitations that should not be ignored. Firstly, the patient population included in this study is relatively small, which was mainly caused by the rarity of this disease. Besides, patients that refrained from surgery were excluded from analysis, which may have introduced unwanted selection bias. The median treatment duration is greater in the surgical benefitting group, which may have been a confounding factor. Nonetheless, when this interval was compared to the non-benefitting group, results were insignificant ($p = 0.10$). In addition, the optimal duration of neoadjuvant treatment in GIST patients is still under debate. Tirumani *et al.* showed an optimal and plateau response at 28 weeks and 34 weeks, respectively [25]. Another limitation involves the heterogeneity in imaging acquisition protocols, which makes it difficult to draw conclusions on the efficacy of density measurements. Standardisation and harmonisation of acquisition and reconstruction protocols may aid towards more reproducible studies on the evaluation of quantitative features obtained from CE-CT imaging [26,27]. Unfortunately, research on this topic is still limited.

All in all, in fifteen patients neoadjuvant treatment was ineffective, as no surgical benefit was obtained. Considering the high costs for some of these targeted therapies and the side-effects patients experienced over the course of treatment, appropriate patient selection for neoadjuvant treatment in GIST patients is warranted. Nonetheless, it would be interesting to evaluate whether surgical benefit can be correlated with an improved progression free survival. In this study, surgical benefit was determined by a single surgical oncologist. To better quantify surgical benefit and overcome inter- and intra-observer variability, the authors propose to involve at least two surgical oncologists. Future research should also focus on *early* response prediction methods. In this case, one could use volumetric criteria as a ground truth. The combined use of different imaging modalities, also known as multimodality imaging, might provide more detailed information that can assist in making early image-guided treatment decisions. [^{18}F]FDG-PET may,

for example, be useful in the early response prediction in GISTs, as metabolic changes seem to precede morphological changes in size [28].

5. Conclusion

Size-based criteria (RECIST 1.1 and volumetry) reflect surgical benefit in GIST patients treated with neoadjuvant therapy most accurately when compared to the Choi criteria. Criteria based on tumour density measurements, such as Choi, showed overestimation of response and a poor predictive power for surgical benefit. Future research should focus on the harmonisation of CE-CT acquisition and reconstruction parameters and the development of response prediction models that assist in making early, image-guided treatment decisions.

6. References

1. Mantese, G. Gastrointestinal stromal tumor: epidemiology, diagnosis, and treatment. *Curr Opin Gastroenterol* 2019, *35*, 555-559, doi:10.1097/mog.0000000000000584.
2. Søreide, K.; Sandvik, O.M.; Søreide, J.A.; Giljaca, V.; Jureckova, A.; Bulusu, V.R. Global epidemiology of gastrointestinal stromal tumours (GIST): A systematic review of population-based cohort studies. *Cancer Epidemiol* 2016, *40*, 39-46, doi:10.1016/j.canep.2015.10.031.
3. Verweij, J.; Casali, P.G.; Zalcberg, J.; LeCesne, A.; Reichardt, P.; Blay, J.Y.; Issels, R.; van Oosterom, A.; Hogendoorn, P.C.; Van Glabbeke, M.; et al. Progression-free survival in gastrointestinal stromal tumours with high-dose imatinib: randomised trial. *Lancet* 2004, *364*, 1127-1134, doi:10.1016/s0140-6736(04)17098-0.
4. Reichardt, P. The Story of Imatinib in GIST - a Journey through the Development of a Targeted Therapy. *Oncol Res Treat* 2018, *41*, 472-477, doi:10.1159/000487511.
5. Lopes, L.F.; Bacchi, C.E. Imatinib treatment for gastrointestinal stromal tumour (GIST). *J Cell Mol Med* 2010, *14*, 42-50, doi:10.1111/j.1582-4934.2009.00983.x.
6. Reynoso, D.; Trent, J.C. Neoadjuvant and adjuvant imatinib treatment in gastrointestinal stromal tumor: current status and recent developments. *Curr Opin Oncol* 2010, *22*, 330-335, doi:10.1097/CCO.0b013e32833aaaad.
7. Rutkowski, P.; Gronchi, A.; Hohenberger, P.; Bonvalot, S.; Schöffski, P.; Bauer, S.; Fumagalli, E.; Nyckowski, P.; Nguyen, B.P.; Kerst, J.M.; et al. Neoadjuvant imatinib in locally advanced gastrointestinal stromal tumors (GIST): the EORTC STBSG experience. *Ann Surg Oncol* 2013, *20*, 2937-2943, doi:10.1245/s10434-013-3013-7.
8. Casali, P.G.; Blay, J.Y.; Abecassis, N.; Bajpai, J.; Bauer, S.; Biagini, R.; Bielack, S.; Bonvalot, S.; Boukovinas, I.; Bovee, J.; et al. Gastrointestinal stromal tumours: ESMO-EURACAN-GENTURIS Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2022, *33*, 20-33, doi:10.1016/j.annonc.2021.09.005.
9. Therasse, P.; Arbuck, S.G.; Eisenhauer, E.A.; Wanders, J.; Kaplan, R.S.; Rubinstein, L.; Verweij, J.; Van Glabbeke, M.; van Oosterom, A.T.; Christian, M.C.; et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst* 2000, *92*, 205-216, doi:10.1093/jnci/92.3.205.
10. Choi, H.; Charnsangavej, C.; de Castro Faria, S.; Tamm, E.P.; Benjamin, R.S.; Johnson, M.M.; Macapinlac, H.A.; Podoloff, D.A. CT evaluation of the response of gastrointestinal stromal tumors after imatinib mesylate treatment: a quantitative analysis correlated with FDG PET findings. *AJR Am J Roentgenol* 2004, *183*, 1619-1628, doi:10.2214/ajr.183.6.01831619.
11. Cohen, N.A.; Kim, T.S.; DeMatteo, R.P. Principles of Kinase Inhibitor Therapy for Solid Tumors. *Ann Surg* 2017, *265*, 311-319, doi:10.1097/sla.0000000000001740.
12. Choi, H.; Charnsangavej, C.; Faria, S.C.; Macapinlac, H.A.; Burgess, M.A.; Patel, S.R.; Chen, L.L.; Podoloff, D.A.; Benjamin, R.S. Correlation of computed tomography and positron emission tomography in patients with metastatic gastrointestinal stromal tumor treated at a single institution with imatinib mesylate: proposal of new computed tomography response criteria. *J Clin Oncol* 2007, *25*, 1753-1759, doi:10.1200/jco.2006.07.3049.
13. Faivre, S.; Demetri, G.; Sargent, W.; Raymond, E. Molecular basis for sunitinib efficacy and future clinical development. *Nat Rev Drug Discov* 2007, *6*, 734-745, doi:10.1038/nrd2380.
14. Jin, T.; Nakatani, H.; Taguchi, T.; Nakano, T.; Okabayashi, T.; Sugimoto, T.; Kobayashi, M.; Araki, K. STI571 (Glivec) suppresses the expression of vascular endothelial growth factor in the gastrointestinal stromal tumor cell line, GIST-T1. *World J Gastroenterol* 2006, *12*, 703-708, doi:10.3748/wjg.v12.i5.703.
15. Dimitrakopoulou-Strauss, A.; Ronellenfitsch, U.; Cheng, C.; Pan, L.; Sachpekidis, C.; Hohenberger, P.; Henzler, T. Imaging therapy response of gastrointestinal stromal tumors (GIST) with FDG PET, CT and MRI: a systematic review. *Clin Transl Imaging* 2017, *5*, 183-197, doi:10.1007/s40336-017-0229-8.

16. Fedorov, A.; Beichel, R.; Kalpathy-Cramer, J.; Finet, J.; Fillion-Robin, J.C.; Pujol, S.; Bauer, C.; Jennings, D.; Fennessy, F.; Sonka, M.; et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging* 2012, *30*, 1323-1341, doi:10.1016/j.mri.2012.05.001.
17. Padhani, A.R.; Ollivier, L. The RECIST 1.1 (Response Evaluation Criteria in Solid Tumors) criteria: implications for diagnostic radiologists. *Br J Radiol* 2001, *74*, 983-986, doi:10.1259/bjr.74.887.740983.
18. van Griethuysen, J.J.M.; Fedorov, A.; Parmar, C.; Hosny, A.; Aucoin, N.; Narayan, V.; Beets-Tan, R.G.H.; Fillion-Robin, J.C.; Pieper, S.; Aerts, H. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* 2017, *77*, e104-e107, doi:10.1158/0008-5472.Can-17-0339.
19. Mackin, D.; Fave, X.; Zhang, L.; Fried, D.; Yang, J.; Taylor, B.; Rodriguez-Rivera, E.; Dodge, C.; Jones, A.K.; Court, L. Measuring Computed Tomography Scanner Variability of Radiomics Features. *Invest Radiol* 2015, *50*, 757-765, doi:10.1097/rli.0000000000000180.
20. Berenguer, R.; Pastor-Juan, M.D.R.; Canales-Vázquez, J.; Castro-García, M.; Villas, M.V.; Mansilla Legorburo, F.; Sabater, S. Radiomics of CT Features May Be Nonreproducible and Redundant: Influence of CT Acquisition Parameters. *Radiology* 2018, *288*, 407-415, doi:10.1148/radiol.2018172361.
21. Ligero, M.; Jordi-Ollero, O.; Bernatowicz, K.; Garcia-Ruiz, A.; Delgado-Muñoz, E.; Leiva, D.; Mast, R.; Suarez, C.; Sala-Llonch, R.; Calvo, N.; et al. Minimizing acquisition-related radiomics variability by image resampling and batch effect correction to allow for large-scale data analysis. *Eur Radiol* 2021, *31*, 1460-1470, doi:10.1007/s00330-020-07174-0.
22. Chen, Y.; Yang, Y.; Wang, C.; Shi, Y. Radiographic response to neoadjuvant therapy and its impact on scope of surgery and prognosis in stage IIB/III soft tissue sarcomas. *BMC Cancer* 2013, *13*, 591, doi:10.1186/1471-2407-13-591.
23. Schiavon, G.; Ruggiero, A.; Schöffski, P.; van der Holt, B.; Bekers, D.J.; Eechoute, K.; Vandecaveye, V.; Krestin, G.P.; Verweij, J.; Sleijfer, S.; et al. Tumor volume as an alternative response measurement for imatinib treated GIST patients. *PLoS One* 2012, *7*, e48372, doi:10.1371/journal.pone.0048372.
24. Bauer, S.; Hartmann, J.T.; de Wit, M.; Lang, H.; Grabellus, F.; Antoch, G.; Niebel, W.; Erhard, J.; Ebeling, P.; Zeth, M.; et al. Resection of residual disease in patients with metastatic gastrointestinal stromal tumors responding to treatment with imatinib. *Int J Cancer* 2005, *117*, 316-325, doi:10.1002/ijc.21164.
25. Tirumani, S.H.; Shinagare, A.B.; Jagannathan, J.P.; Krajewski, K.M.; Ramaiya, N.H.; Raut, C.P. Radiologic assessment of earliest, best, and plateau response of gastrointestinal stromal tumors to neoadjuvant imatinib prior to successful surgical resection. *Eur J Surg Oncol* 2014, *40*, 420-428, doi:10.1016/j.ejso.2013.10.021.
26. Sachs, P.B.; Hunt, K.; Mansoubi, F.; Borgstede, J. CT and MR Protocol Standardization Across a Large Health System: Providing a Consistent Radiologist, Patient, and Referring Provider Experience. *J Digit Imaging* 2017, *30*, 11-16, doi:10.1007/s10278-016-9895-8.
27. Mali, S.A.; Ibrahim, A.; Woodruff, H.C.; Andrearczyk, V.; Müller, H.; Primakov, S.; Salahuddin, Z.; Chatterjee, A.; Lambin, P. Making Radiomics More Reproducible across Scanner and Imaging Protocol Variations: A Review of Harmonization Methods. *J Pers Med* 2021, *11*, doi:10.3390/jpm11090842.
28. Weeda, Y.A.; Kalisvaart, G.M.; van Velden, F.H.P.; Gelderblom, H.; van der Molen, A.J.; Bovee, J.; van der Hage, J.A.; Grootjans, W.; de Geus-Oei, L.F. Early Prediction and Monitoring of Treatment Response in Gastrointestinal Stromal Tumors by Means of Imaging: A Systematic Review. *Diagnostics (Basel)* 2022, *12*, doi:10.3390/diagnostics12112722.

3

Added value of CE-CT and [^{18}F]FDG-PET radiomics in predicting volumetric response in gastrointestinal stromal tumour patients treated with neoadjuvant intent

Added value of CE-CT and [^{18}F]FDG-PET radiomics in predicting volumetric response in gastrointestinal stromal tumour patients treated with neoadjuvant intent

Ylva A. Weeda[1], Gijsbert M. Kalisvaart[1], Floris H.P. van Velden[1], Aart J. van der Molen[1], Joost J.H. Roelofs[1], Jos A. van der Hage[2], Lioe-Fee de Geus-Oei[1,3,4], Willem Grootjans[1]

[1]Department of Radiology, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands.
[2]Department of Surgical Oncology, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands.
[3]Biomedical Photonic Imaging Group, University of Twente, 7522 NB Enschede, The Netherlands
[4]Department of Radiation Science & Technology, Technical University of Delft, 2629 JB Delft, The Netherlands

Abstract: Predicting response in gastrointestinal stromal tumours (GISTs) is complex, particularly due to the biological intra- and intertumoral heterogeneity. Currently only radiological criteria can be used to determine response after treatment induction. The aim of this study is to investigate whether radiomic models based on diagnostic contrast-enhanced computed tomography (CE-CT) and 2-deoxy-2-[^{18}F]fluoro-D-glucose positron emission tomography ([^{18}F]FDG-PET) imaging can be used to predict volume reduction in patients receiving neo-adjuvant tyrosine kinase inhibitor (TKI) treatment. Forty non-metastatic GIST patients treated with neoadjuvant intent and monitored with consecutive portal venous CE-CT scans, were included. In 18 of these patients baseline [^{18}F]FDG PET/CT scans were available and were used for analysis. Response (> 10 weeks into treatment) was defined by a change in tumour volume of more than 66%. Tumours were segmented manually and radiomics features (first-order, shape and texture features) were extracted. Four radiomic classifiers were trained with both CE-CT and [^{18}F]FDG PET/CT imaging features and tested by means of a stratified nested cross-validation approach with fifty iterations. Feature selection and hyperparameter tuning methods were introduced in the internal validation. The performance of a radiological scoring system (response evaluation criteria in solid tumours 1.1 (RECIST 1.1)), determined on baseline, first response follow-up and last preoperative portal venous CE-CT response scans, was used to evaluate the added value of the radiomic models . The mean area under the curve (AUC) values for the radiomic models were all below 0.50. Overall performance could be improved by increasing the sample size and concurrent harmonisation of imaging acquisition protocols. On the other hand, manual RECIST 1.1 measurements predicted volumetric response with an AUC of 0.89 after the first response follow-up and 0.93 after the last preoperative response scan. RECIST 1.1 measurements could aid in early surgical decision making and improve clinical management of GISTs.

1. Introduction

Gastrointestinal stromal tumours (GISTs) are rare mesenchymal neoplasms affecting the entire gastrointestinal tract, but most commonly the stomach and small intestine [1]. Neoadjuvant tyrosine kinase inhibitor (TKI) therapy is administered in selected patients to attain size reductions of the primary tumour and improve chances of complete resection [2,3]. Nevertheless, defining response in GIST patients is challenging. This is mainly attributed to intra- and intertumoral heterogeneity and the fact that there is no definitive correlation between pathological response and clinical implications. The percentage of remaining viable tumour cells in surgical specimens is often used in other mesenchymal tumours to define response. However, this histopathological criterion is not correlated with overall survival in GIST patients [4].

Currently, tumour response is typically determined through radiographic criteria derived from consecutive contrast-enhanced computed tomography (CE-CT scans) that measure changes in tumour size and density over the course of neoadjuvant treatment [5,6]. Positron emission tomography combined with CT and the radiotracer 2-deoxy-2-[^{18}F]fluoro-D-glucose ([^{18}F]FDG-PET/CT) is an imaging modality used to visualise changes in metabolic behaviour. Although metabolic changes seem to precede morphological changes in size, the imaging modality is not routinely used for monitoring because of higher costs [7-11]. In the previous chapter, size-based criteria (response evaluation criteria in solid tumours (RECIST 1.1) and volumetry), were shown to reflect the obtained surgical benefit, compared to density measurements as presented by the Choi criteria (see Chapter 2). RECIST 1.1 and volumetric criteria are also less prone to scanner and imaging protocol variabilities. Nonetheless, responder status can only be determined *during* or *after* treatment induction.

Therefore, there is a need for upfront prediction of response in order to facilitate appropriate patient selection for neoadjuvant TKI therapy at baseline and protect patients from futile treatment, concurrent side-effects and healthcare costs. Radiomics includes the extraction and analysis of quantitative features from medical imaging data to provide additional information on tumour biology [12,13]. The aim of this study is to develop a radiomic model using baseline CE-CT and [^{18}F]FDG-PET imaging features and evaluate its performance to predict volumetric response. CE-CT radiomics has been applied for risk stratification and the prediction of mutational status, proliferative activity and prognosis in GIST patients [11]. To the author's knowledge, this is the first CE-CT and [^{18}F]FDG-PET radiomic study attempting to predict radiological response in GIST patients treated with neoadjuvant intent.

2. Methods

2.1 Patients

In this two-centre radiomic study, patients with a confirmed non-metastatic GIST diagnosis and consecutive reference for neoadjuvant TKI treatment and surgical resection were retrospectively identified at the Leiden University Medical Center (LUMC) and the Netherlands Cancer Institute-Antoni van Leeuwenhoek (NKI-AvL). Patients were monitored with portal venous phase CE-CT and had baseline CE-CT and [^{18}F]FDG PET/CT scans. This study was executed under the Dutch GIST consortium and patient consent was waived by the local ethical board, as data were retrospectively collected and pseudo-anonymised.

2.2 Responder status

The ground truth on responder status was determined by a tumour volume reduction above 66%, as it best reflected the obtained surgical benefit after neo-adjuvant TKI-treatment (see Chapter 2). Tumours were manually delineated on the baseline and the last preoperative portal venous response CT scans. Baseline scans were made within three months prior to TKI treatment induction. The percentage change in tumour

volume was computed for each individual patient. Volumetric response was defined according to specific thresholds, which introduced four types of response; complete response (clearance of entire tumour), partial response (≥ 66% reduction in tumour volume), progressive disease (≥ 73% increase in tumour volume) and stable disease (neither progressive disease nor partial response). These thresholds for volumetric measurements were derived from the current radiographic Response Evaluation Criteria in Solid Tumour 1.1 (RECIST 1.1) in which a partial response is defined as a 30% reduction in transaxial tumour diameter. According to the square cube law, an isotropic 30% reduction in tumour diameter translates to a 66% reduction in tumour volume. Good responders were considered to have complete or partial response, while patients with progressive and stable disease were defined as poor responders.

2.3 Image segmentation

Portal venous phase CE-CT and [^{18}F]FDG-PET/CT performed prior to TKI administration were collected for each individual patient. Images with the smallest slice thickness were selected for the segmentation process. Tumour segmentation was performed by a technical medicine master student (YAW), using the 3D slicer (version 5.0.3) image computing platform. Lesions were manually delineated in the axial plane and automatically filled using slice-to-slice interpolation to create binary three-dimensional (3D) tumour masks. The concurrent low-dose non-contrast-enhanced CT images were used to guide the delineation of the tumour on [^{18}F]FDG-PET imaging. The entire segmentation process was supervised by an expert abdominal radiologist. To assess inter-observer variability, a representative subset of images (five CE-CT scans and three [^{18}F]FDG-PET scans) were delineated by a second observer, namely a medical PhD student (GMK).

2.4 Feature extraction and selection

For each patient, first order, shape and texture features were extracted from the three-dimensional (3D) tumour volume, using Python (version 3.7) and the Pyradiomics software (version 3.0.1) [14]. Highly correlated features (Pearson correlation > 0.95) and non-distinctive features (zero variance) were omitted from the radiomic analysis. The highly correlated features were dropped from the upper triangle of the correlation matrix. The standard scalar (z-score normalisation) method was used to scale the feature data such that the mean and standard deviation were equal to one and zero, respectively. Features extracted from portal venous CE-CT imaging were placed inside dataset A and a combination of both CE-CT and [^{18}F]FDG-PET features were placed in dataset B. To reduce feature dimensionality and improve generalisability of the model, the least absolute shrinkage and selection operator (LASSO) was used for feature selection. The most favourable value for λ (parameter controlling regularisation strength) was selected by averaging the outcomes of a LASSO cross-validation (5-fold, 20 repeats and 1000 iterations).

2.5 Model development

Four machine learning classifiers were selected to predict responder status in GIST patients, including Logistic Regression, k-Nearest Neighbours (k-NN), Random Forest (RF) and Support Vector Machine (SVM). To evaluate the predictive power of the extracted radiomic features in classifying poor and good responders, the described classifiers were trained and tested. The feature values and responder labels were divided into an 80% training set and 20% test set, using a stratified five-fold cross validation with fifty iterations. The hyperparameter tuning was executed within an internal five-fold cross-validation and fifteen iterations. The F1-score (Equation 1.1) was used as performance metric. Inside the internal validation, the previously described standard scaler and LASSO feature selection method were also introduced.

$$F1\text{-}score = 2 * \frac{precision * recall}{precision + recall} \quad (1.1)$$

2.6 Model evaluation

The two datasets (A and B) were used to develop two models and they were all assessed through the following outcome measures: F1-score, accuracy, sensitivity and specificity. Subsequently, the Receiving Operator Characteristic (ROC) curves were plotted per model, including the mean Area Under the Curve (AUC). Manual RECIST 1.1 measurements were executed by a technical medicine student (YAW) on portal venous CE-CT imaging made at baseline, after the first follow-up and prior to surgery [6]. The RECIST 1.1 measurements were supervised by a professional radiographer with four years of experience in RECIST 1.1 reporting (JJHR). Good responders were considered to have complete or partial response, while patients with stable or progressive disease were classified as poor responders. The findings on RECIST 1.1 measurements were used for gold standard comparison of the radiomic models to evaluate its added clinical value. The evaluation pipeline is graphically displayed in Fig 1. All experiments were executed using an Intel(R) Core(TM) i3-10100 CPU @ 3.60GHz CPU with 64-bit Windows operating system.

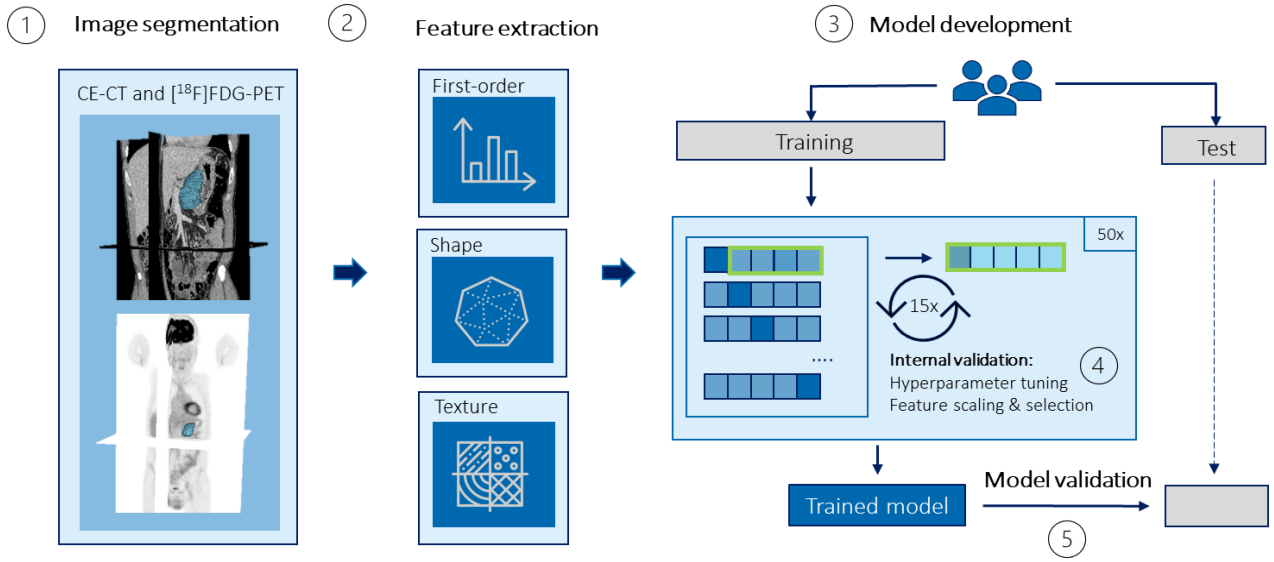

Fig 1. Schematic representation of the radiomics evaluation pipeline with a nested-cross validation approach. 1) CE-CT and [^{18}F]FDG-PET images were processed (resampling and intensity discretisation) and the GISTs were manually delineated to create three-dimensional tumour volumes. 2) First-order, shape and texture radiomic features were extracted from the tumour volumes. 3) These features were split into a training and a test set and were used to train predefined classifiers. 4) Hyperparameter tuning was executed through internal validation, where feature selection and scaling methods were also introduced. 5) The final trained models are validated on the test set through F1-score, accuracy, sensitivity and specificity.

2.7 Statistical analysis

Differences between the good and poor responders at baseline were assessed using Python (version 3.7) software. Mann-Whitney U test/independent t-test and chi-square tests were performed for numerical data and categorical data, respectively. A p-value below 0.05 was considered statistically significant. Inter-observer variability in the manual segmentation process was evaluated using the intra-class correlation coefficient (ICC) and the Dice Similarity Score (DSC).

3. Preliminary results

3.1 Patient characteristics

This study shows the preliminary results of the patients that were included in the LUMC, which involved a total of 40 non-metastatic GIST patients. All patients received portal venous CE-CT imaging at the time of diagnosis and 18 of these patients received additional [^{18}F]FDG-PET imaging. All patients received a 400 mg daily imatinib mesylate dosage, except for one patient who received a daily avapritinib dosage of 300 mg. This was administered because of the presence of a PDGFRA exon 18 (D842V) mutation. TKI therapy was terminated just before surgery, which resulted in a median treatment duration of 294.5 days. Based on the aforementioned volumetric response criteria, 20 patients were classified as good responders and 20 patients as poor responders. In the [^{18}F]FDG-PET group, ten patients were considered to have a good response, while eight patients had a poor response. The treatment duration was significantly higher in the group of good responders in both the CE-CT and [^{18}F]FDG-PET imaging group (< 0.01). RECIST 1.1 measurements were executed on the first response CE-CT scan (median of 2.8 months, IQR 2.1-3.3) and the final preoperative response CE-CT scan (median of 7.7 months, IQR 3.8-10.7). The first follow-up scan was also the last response scan before surgical resection in nine patients. All patient characteristics have been summarised in Table 1.

Patient characteristics						
	Portal venous CE-CT			[^{18}F]FDG-PET		
	Poor (n = 20)	Good (n = 20)	p-value	Poor (n = 8)	Good (n = 10)	p-value
Age (years)	65.5 (54.8 - 72.0)	61.5 (46.0 - 65.5)	0.24	69.0 (61.0, 74.0)	61.5 (57.0m 65.0)	0.15
Sex			0.30			1.0
Female	4 (20.0%)	8 (40.0%)		2 (25.0%)	3 (30.0%)	
Male	16 (80.0%)	12 (60.0%)		6 (75.0%)	7 (70.0%)	
Tumour location			1.0			0.56
Stomach	14 (70.0%)	16 (80.0%)	*Upper vs.*	6 (75.0%)	8 (80.0%)	*Upper vs.*
Duodenum	4 (20.0%)	1 (5.0 %)	*lower GI tract*	0 (0.0%)	1 (10.0 %)	*lower GI tract*
Small intestine	1 (5.0 %)	1 (5.0 %)		1 (12.5 %)	1 (10.0 %)	
Rectum	1 (5.0 %)	2 (10.0 %)		1 (12.5 %)	0 (0.0%)	
Largest primary tumour size (mm)	104.0 (70.8, 154.2)	105.5 (81.5 - 218.5)	0.41	93.5 (74.2, 154.2)	96.0 (79.2, 124.0)	0.89
Mitotic index			0.66			1.0
≤ 5 per 5 mm^2	8 (40.0%)	9 (45.0%)	*High vs. low*	3 (37.5%)	4 (40.0%)	*High vs. low*
> 5 per 5 mm^2	2 (10.0%)	4 (20.0%)	*mitotic index*	2 (25.0%)	2 (20.0%)	*mitotic index*
Not reported	10 (50.0%)	7 (35.0%)		3 (37.5%)	4 (40.0%)	
Risk stratifciation			1.0			1.0
Low	5 (25.0%)	2 (10.0%)	*High vs.*	2 (25.0%)	2 (20.0%)	*High vs.*
Intermediate	2 (10.0%)	6 (30.0%)	*Intermediate*	1 (12.5%)	2 (20.0%)	*Intermediate*
High	3 (15.0%)	5 (25.0%)	*and low risk*	2 (25.0%)	2 (20.0%)	*and low risk*
Not reported	10 (50.0%)	7 (17.5%)		3 (37.5%)	4 (40.0%)	
Mutational status			0.08			0.12
KIT exon 11	11 (55.0%)	17 (85.0%)	*KIT exon 11*	4 (50.0%)	9 (90.0%)	*KIT exon 11*
KIT exon 9	1 (5.0 %)	1 (5.0%)	*vs. other*	0 (0.0%)	0 (0.0%)	*vs. other*
PDGFRA exon 14	1 (5.0 %)	0 (0.0%)	*mutations*	0 (0.0%)	0 (0.0%)	*mutations*
PDGFRA exon 18	5 (25.0%)	2 (10.0,%)		3 (37.5%)	1 (10.0,%)	
Wildtype	2 (0.0%)	0 (0.0%)		1 (12.5%)	0 (0.0%)	
Treatment duration (days)	200.0 (130.2 - 289.8)	348.0 (297.8 - 395.8)	< 0.01*	198.0 (130.2 - 260.5)	323.5 (270.0 - 385.8)	< 0.01*

Table 1. Baseline characteristics of the included non-metastatic GIST patients, stratified by responder status and imaging modality. Responder status was determined through volumetric response criteria. Numerical data is presented as a median value, along with the concurrent interquartile range. Risk stratification was assessed using the Miettinen classification system involving mitotic index, tumour size and location as prognostic factors. Tumour size was in this case defined by the largest tumour dimension. P-values for numerical data were calculated through the Mann-Whitney and the Fisher's exact test was used for categorical data. (GI = gastrointestinal, PDGFRA = platelet-derived growth factor alpha, KIT = receptor tyrosine kinase). * Upper gastrointestinal tract (stomach and duodenum), lower gastrointestinal tract (small intestine and rectum). † Non-reported cases were omitted from statistical analysis.

3.2 Image acquisition and processing

Images were acquired at different hospitals, introducing heterogeneity in image acquisition parameters. The CE-CT scans originated from four different manufacturers, including Siemens Healthineers, GE Healthcare, Canon Medical Systems (previously known as Toshiba Medical Systems). CE-CT imaging was initiated around 35-40 seconds after intravenous administration of an iodinated contrast agent to facilitate

enhancement in the portal venous phase. Median scanning parameters were as follows: 120 peak tube voltage (kVp), 1.0 mm slice thickness and 0.8 pixel spacing. Baseline [^{18}F]FDG-PET images were acquired by scanners from three manufacturers (GE Healthcare, Philips Healthcare and Siemens Healthineers). In 14 out 18 patients, a standard acquisition and reconstruction protocol was adhered, which is in accordance with the European Association of Nuclear Medicine (EANM) Research Ltd (EARL) guidelines for FDG-PET tumour imaging [15]. Image acquisition was initiated around 50-69 minutes after intra-venous bolus administration of [^{18}F]FDG. Images had a median slice thickness and pixel spacing of 4.0 mm. The maximum pixel size and slice thickness for CE-CT imaging (0.9 mm and 5.0 mm) and [^{18}F]FDG-PET imaging (4.1 mm and 5.0 mm) were used for resampling to create isotropic voxels. Fixed bin size intensity discretisation was implemented for both CE-CT (bin width of 25 Hounsfield Unit) and [^{18}F]FDG-PET imaging (bin width of 0.25 standardised uptake value). An overview of image acquisition parameters for both modalities can be found in Supplementary Materials (Table S1).

3.3 Assessment of intra-observer variability

The subset of images (five CE-CT scans and three [^{18}F]FDG-PET scans) that were delineated by two observers had a mean DSC of 0.92 (± 0.03). The mean tumour volumes of patients who received both CE-CT and [^{18}F]FDG-PET imaging measured $8.49 \bullet 10^2$ cm^3 (± $14.6 \bullet 10^2$) and $8.16 \bullet 10^2$ cm^3 (± $14.4 \bullet 10^2$), respectively. The agreement across the segmentations and between the two imaging modalities were considered acceptable.

3.4 Feature extraction and selection

In total 107 features were extracted from each individual image, involving 18 first-order, 14 shape and 75 texture features. After removal of highly correlated and non-distinctive features, dataset A contained 46 CE-CT features (12 first-order, 7 shape and 27 texture features) for 40 patients, while dataset B contained 90 features (25 first-order, 11 shape and 54 texture features) for both modalities and 18 patients. The mean ICC for the extracted features across the segmentations was 0.88 (± 0.27) and 0.98 (± 0.07) for CE-CT and [^{18}F]FDG-PET imaging, respectively.

A five-fold LASSO cross-validation was used to find the most favourable λ for dataset A (0.099) and B (0.115). Cluster prominence, energy and 90 percentile were the most commonly selected features from the CE-CT dataset during training. In dataset B, [^{18}F]FDG-PET features (minimum, 10 percentile and flatness) were selected in greater frequency when compared to CE-CT features. Of note, the minimum value on [^{18}F]FDG-PET imaging was significantly higher in the responding group. Bar plots on selection frequency can be found in Supplementary Materials (Fig S1-2).

3.5 Model development and validation

During training, classifiers showed improvement in F1-score when hyperparameter tuning and feature selection were introduced. No performance difference between baseline and hyperparameter tuning was observed in the LR classifier, as the default parameters were used throughout the entire training process. Comparable F1-scores were computed for the four classifiers, and it was therefore decided to include all classifiers for final evaluation. The results from the training process are summarised in Supplementary Materials (Table S2). Based on the manual RECIST 1.1 measurements, a good response was observed in 18 patients and a poor response in 22 patients. In the subset of patients that received additional [^{18}F]FDG-PET imaging at baseline, 7 patients were classified as good responders and 11 as poor responders. The performance of the radiomic models in predicting response is shown in Table 2. The KNN and RF classifiers score the highest accuracy, specificity, F1-score and AUC for dataset A and B respectively. However, AUCs are all below 0.50 indicating an even lower performance compared to a random model. The gold standard RECIST 1.1 measurements, predicted volumetric response with high accuracy after the first response follow-up (80.0%) and prior to surgery (85.0%) in all forty patients. The RECIST 1.1 measurements performed after

the first response scan yield a high sensitivity of 95.0%, however at a lower specificity (65.0%). The specificity increased when the RECIST 1.1 measurements were performed after a longer period of time (median interval of 7.7 months). ROC curves for the models with the highest AUC values, along with the gold standards, are visualised in Fig. 2.

Performance scores										
	Dataset A (CE-CT)					Dataset B (CE-CT + [^{18}F]FDG-PET)				
Classifiers	Accuracy	Sensitivity	Specificity	F1-score	AUC	Accuracy	Sensitivity	Specificity	F1-score	AUC
LR	0.43 (0.14)	0.46 (0.24)	0.39 (0.24)	0.38 (0.20)	0.41 (0.17)	0.33 (0.22)	0.24 (0.36)	0.41 (0.34)	0.36 (0.28)	0.22 (0.25)
KNN	0.45 (0.15)	0.46 (0.25)	0.43 (0.25)	0.41 (0.21)	0.44 (0.16)	0.36 (0.20)	0.21 (0.33)	0.48 (0.32)	0.42 (0.26)	0.31 (0.22)
RF	0.39 (0.15)	0.39 (0.24)	0.40 (0.26)	0.37 (0.21)	0.35 (0.17)	0.48 (0.24)	0.30 (0.38)	0.63 (0.36)	0.54 (0.28)	0.42 (0.34)
SVM	0.44 (0.14)	0.45 (0.26)	0.43 (0.27)	0.40 (0.27)	0.43 (0.18)	0.38 (0.21)	0.24 (0.37)	0.50 (0.35)	042 (0.27)	0.32 (0.31)
Gold standard	Accuracy	Sensitivity	Specificity	F1-score	AUC	Accuracy	Sensitivity	Specificity	F1-score	AUC
RECIST first response	0.80	0.95	0.65	0.76	0.89	0.78	1.00	0.60	0.75	0.78
RECIST final response	0.85	0.90	0.80	0.84	0.93	0.72	0.88	0.60	0.71	0.84

Table 2. Performance scores for four machine learning classifiers using dataset A (portal venous CE-CT imaging features) and dataset B (combination of CE-CT and [^{18}F]FDG-PET imaging features) to predict volumetric response in non-metastatic GIST patients. Manual RECIST measurements, acquired from the first and last portal venous CE-CT response scans, were used for gold standard comparison. The two models with the highest AUC value are framed with a green box (KNN for dataset A and RF for dataset B). LR = logistic regression, KNN = k-nearest neighbour, RF = random forest, SVM = support vector machine, AUC = area under the curve.

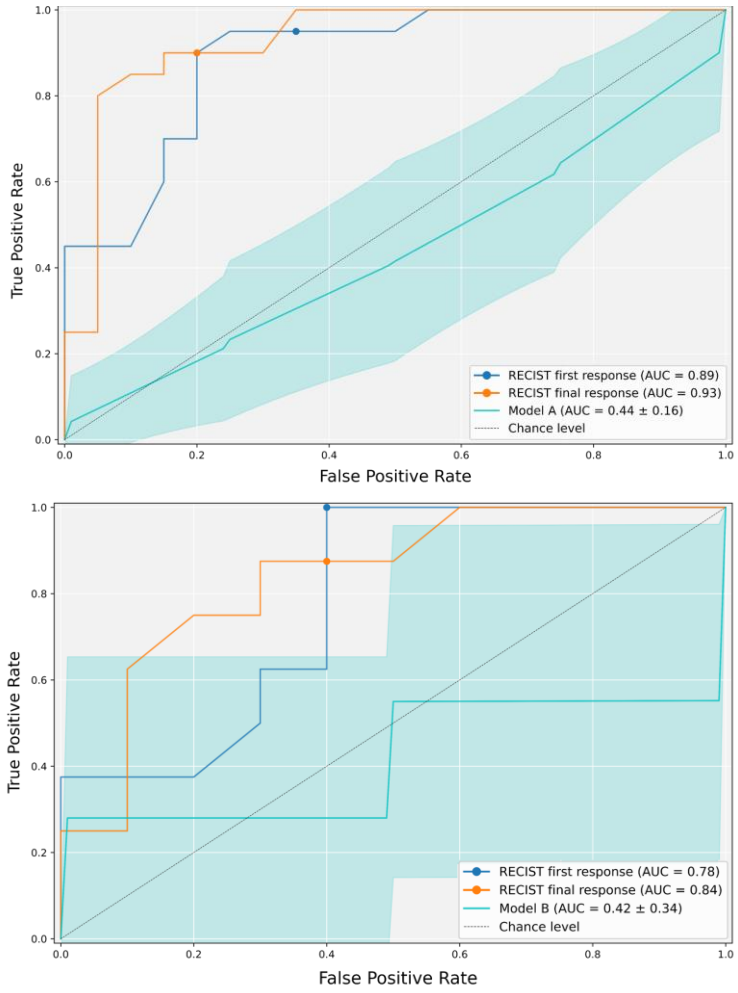

Fig. 2 Receiver operating characteristic curves for model A (upper) and model B (lower) used to predict volumetric response in non-metastatic GIST patients. The results are obtained from a five-fold cross validation with fifty iterations as presented in the methods section. The blue area demonstrates the computed standard deviations. The manual RECIST measurements, acquired from the first (blue) and last (orange) response portal venous CE-CT scans, were used for gold standard comparison. The cut-off points (30% tumour diameter reduction) for these standards were visualised by dots. The chance level is displayed by the dotted black line.

4. Discussion

This study shows the preliminary results of a CE-CT and [^{18}F]FDG-PET radiomic pipeline used for upfront prediction of volumetric response in non-metastatic GIST patients treated with neoadjuvant intent. Results should be interpreted carefully given the very small size of the patient cohort analysed.

Many articles have been published on the use of radiomic and deep learning models for response prediction on baseline CE-CT imaging in the clinical management of GIST patients. High performance scores were stated for models predicting recurrence free survival and risk stratifications, while mutational status remained difficult to predict with variable AUC values [11,16-23]. To the author's knowledge, this is the first CE-CT and [^{18}F]FDG-PET radiomic study attempting to predict volumetric response in GIST patients treated with neoadjuvant intent. The radiomic models presented in this study overfitted on the training data, causing an overall low performance scores when applied to the test set. Lack of generalisation and poor performance of the trained models in this study could be attributed to the limited number of samples (especially in dataset B combining CE-CT and [^{18}F]FDG-PET imaging features) and the presence of heterogeneity in the imaging acquisition protocols. However, the true cause can only be determined through additional data inclusion.

[^{18}F]FDG-PET imaging is considered a suitable method to monitor treatment response, as metabolic changes can be observed within seven days of treatment and precede measurable changes in size [8,10,24-26]. In this study, the minimum value on [^{18}F]FDG-PET imaging was frequently selected during the training process. This value was significantly higher in the responding group, which could suggest a more uniform tracer uptake in TKI responders at baseline. Miyake et al. found ring-shaped [^{18}F]FDG-PET uptake to be an independent adverse prognostic factor for postoperative recurrence. This uptake pattern was presumably related to the presence of coagulative tumour cell and hyaline necrosis at the centre of the lesion [27]. We therefore hypothesise uniform uptake to be associated with less aggressive tumour behaviour and better response to TKI therapy. Knowledge on underlying tumour biology aids in deterministic feature selection and improved response prediction. However future research is warranted to confirm our hypothesis. Another advantage of [^{18}F]FDG-PET imaging includes the use of more standardised acquisition and reconstruction protocols (EARL initiative) throughout different hospitals. The use of EARL protocols helps to improve quantitative accuracy of PET images. As a consequence, there is less chance of data heterogeneity. Nonetheless, due to the limited number of patients receiving diagnostic [^{18}F]FDG-PET imaging in the LUMC, this study could not provide comprehensive results on its added value in clinical practice.

Literature reports underestimations of response to TKI therapy in GIST patients when using unidimensional RECIST 1.1 measurements [11]. Nevertheless, RECIST 1.1 measurements could accurately predict volumetric response in this non-metastatic GIST patient population. Volumetric response could already be determined after a median follow-up of about 11 weeks in 32 patients. In a study by Tirumani et al, a median earliest response (time when partial response was observed) of 16 weeks was found [28]. The difference in interval may be caused by variations in patient population, as our current study included a smaller amount of rectal GISTs (8 vs. 3) and median primary tumour dimensions were higher (7.2 cm vs. 10.4 cm). Surgery is performed earlier in case of minimal tumour shrinkage, and we can therefore not rule out that substantial tumour shrinkage could still occur over longer TKI treatment durations. However, this does not take away from the fact that a poor response according to RECIST 1.1 was associated with patients having no surgical benefit after neoadjuvant treatment induction. Even though tumour diameter reductions according to RECIST 1.1 seem promising, the criteria should be executed properly. The tumour diameter should only be measured in the axial plane, which is often forgotten. Nonetheless, current results suggest that RECIST 1.1 measurements may contribute to the early prediction of TKI therapy response in GIST patients without the

requirement to segment the entire tumour (which is labour intensive) and should therefore be considered in clinical practice.

The limitation of the current study is the small size of the patient cohort. Additional imaging data will be added to improve overall performance. An increased sample size also opens up the possibility for Combining Batches (ComBat) harmonisation, to reduce the heterogeneity within the imaging acquisition protocols. In order to increase the radiomics quality score (RQS) of this study, it would be suggested to include calibration statistics and cost-effectiveness analysis. This radiomic study assigns the same weights for all extracted features. Future research should include a more hypothesis-driven radiomic approach, where we select predefined quantitative features from different imaging modalities to reflect tumour intra- and tumoral heterogeneity.

5. Conclusion

The radiomic models presented in this study showed poor performance and can therefore not yet be applied in a clinical setting. However, in current retrospective results, tumour diameter reductions according to RECIST 1.1 after the first response follow-up scan are associated with preoperative volumetric response. These manual measurements could therefore aid towards early surgical decision making and improve clinical management of GISTs. Future research should focus on additional data collection, harmonisation of acquisition protocols and hypothesis-driven radiomics.

6. References

1. Balachandran, V.P.; DeMatteo, R.P. Gastrointestinal stromal tumors: who should get imatinib and for how long? *Adv Surg* 2014, *48*, 165-183, doi:10.1016/j.yasu.2014.05.014.
2. Lopes, L.F.; Bacchi, C.E. Imatinib treatment for gastrointestinal stromal tumour (GIST). *J Cell Mol Med* 2010, *14*, 42-50, doi:10.1111/j.1582-4934.2009.00983.x.
3. Reynoso, D.; Trent, J.C. Neoadjuvant and adjuvant imatinib treatment in gastrointestinal stromal tumor: current status and recent developments. *Curr Opin Oncol* 2010, *22*, 330-335, doi:10.1097/CCO.0b013e32833aaaad.
4. Bauer, S.; Hartmann, J.T.; de Wit, M.; Lang, H.; Grabellus, F.; Antoch, G.; Niebel, W.; Erhard, J.; Ebeling, P.; Zeth, M.; et al. Resection of residual disease in patients with metastatic gastrointestinal stromal tumors responding to treatment with imatinib. *Int J Cancer* 2005, *117*, 316-325, doi:10.1002/ijc.21164.
5. Choi, H.; Charnsangavej, C.; Faria, S.C.; Macapinlac, H.A.; Burgess, M.A.; Patel, S.R.; Chen, L.L.; Podoloff, D.A.; Benjamin, R.S. Correlation of computed tomography and positron emission tomography in patients with metastatic gastrointestinal stromal tumor treated at a single institution with imatinib mesylate: proposal of new computed tomography response criteria. *J Clin Oncol* 2007, *25*, 1753-1759, doi:10.1200/jco.2006.07.3049.
6. Padhani, A.R.; Ollivier, L. The RECIST (Response Evaluation Criteria in Solid Tumors) criteria: implications for diagnostic radiologists. *Br J Radiol* 2001, *74*, 983-986, doi:10.1259/bjr.74.887.740983.
7. Young, H.; Baum, R.; Cremerius, U.; Herholz, K.; Hoekstra, O.; Lammertsma, A.A.; Pruim, J.; Price, P. Measurement of clinical and subclinical tumour response using [18F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET Study Group. *Eur J Cancer* 1999, *35*, 1773-1782, doi:10.1016/s0959-8049(99)00229-4.
8. Gayed, I.; Vu, T.; Iyer, R.; Johnson, M.; Macapinlac, H.; Swanston, N.; Podoloff, D. The role of 18F-FDG PET in staging and early prediction of response to therapy of recurrent gastrointestinal stromal tumors. *J Nucl Med* 2004, *45*, 17-21.
9. Beheshti, M.; Vali, R.; Langsteger, W.; Li, S.; Schima, W.; Dudczak, R. The potential value of F-18 FDG PET in comparison to ct in early prediction of response to Imatinib (STI571) therapy in patients with gastrointestinal stromal tumors. *Iranian Journal of Nuclear Medicine* 2007, *15*, 34-42.
10. Van den Abbeele, A.D.; Gatsonis, C.; de Vries, D.J.; Melenevsky, Y.; Szot-Barnes, A.; Yap, J.T.; Godwin, A.K.; Rink, L.; Huang, M.; Blevins, M.; et al. ACRIN 6665/RTOG 0132 phase II trial of neoadjuvant imatinib mesylate for operable malignant gastrointestinal stromal tumor: monitoring with 18F-FDG PET and correlation with genotype and GLUT4 expression. *J Nucl Med* 2012, *53*, 567-574, doi:10.2967/jnumed.111.094425.
11. Weeda, Y.A.; Kalisvaart, G.M.; van Velden, F.H.P.; Gelderblom, H.; van der Molen, A.J.; Bovee, J.; van der Hage, J.A.; Grootjans, W.; de Geus-Oei, L.F. Early Prediction and Monitoring of Treatment Response in Gastrointestinal Stromal Tumors by Means of Imaging: A Systematic Review. *Diagnostics (Basel)* 2022, *12*, doi:10.3390/diagnostics12112722.
12. Gillies, R.J.; Kinahan, P.E.; Hricak, H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* 2016, *278*, 563-577, doi:10.1148/radiol.2015151169.
13. Lubner, M.G.; Smith, A.D.; Sandrasegaran, K.; Sahani, D.V.; Pickhardt, P.J. CT Texture Analysis: Definitions, Applications, Biologic Correlates, and Challenges. *Radiographics* 2017, *37*, 1483-1503, doi:10.1148/rg.2017170056.
14. van Griethuysen, J.J.M.; Fedorov, A.; Parmar, C.; Hosny, A.; Aucoin, N.; Narayan, V.; Beets-Tan, R.G.H.; Fillion-Robin, J.C.; Pieper, S.; Aerts, H. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* 2017, *77*, e104-e107, doi:10.1158/0008-5472.Can-17-0339.
15. Boellaard, R.; Delgado-Bolton, R.; Oyen, W.J.; Giammarile, F.; Tatsch, K.; Eschner, W.; Verzijlbergen, F.J.; Barrington, S.F.; Pike, L.C.; Weber, W.A.; et al. FDG PET/CT: EANM procedure

guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging* 2015, *42*, 328-354, doi:10.1007/s00259-014-2961-x.
16. Chen, T.; Ning, Z.; Xu, L.; Feng, X.; Han, S.; Roth, H.R.; Xiong, W.; Zhao, X.; Hu, Y.; Liu, H.; et al. Radiomics nomogram for predicting the malignant potential of gastrointestinal stromal tumours preoperatively. *Eur Radiol* 2019, *29*, 1074-1082, doi:10.1007/s00330-018-5629-2.
17. Chen, Z.H.; Xu, L.Y.; Zhang, C.M.; Huang, C.C.; Wang, M.H.; Feng, Z.; Xiong, Y. CT Radiomics Model for Discriminating the Risk Stratification of Gastrointestinal Stromal Tumors: A Multi-Class Classification and Multi-Center Study. *FRONTIERS IN ONCOLOGY* 2021, *11*.
18. Ren, C.; Wang, S.; Zhang, S. Development and validation of a nomogram based on CT images and 3D texture analysis for preoperative prediction of the malignant potential in gastrointestinal stromal tumors. *Cancer Imaging* 2020, *20*, 5, doi:10.1186/s40644-019-0284-7.
19. Wang, M.; Feng, Z.; Zhou, L.; Zhang, L.; Hao, X.; Zhai, J. Computed-Tomography-Based Radiomics Model for Predicting the Malignant Potential of Gastrointestinal Stromal Tumors Preoperatively: A Multi-Classifier and Multicenter Study. *Front Oncol* 2021, *11*, 582847, doi:10.3389/fonc.2021.582847.
20. Zhang, L.; Kang, L.; Li, G.; Zhang, X.; Ren, J.; Shi, Z.; Li, J.; Yu, S. Computed tomography-based radiomics model for discriminating the risk stratification of gastrointestinal stromal tumors. *Radiol Med* 2020, *125*, 465-473, doi:10.1007/s11547-020-01138-6.
21. Zhang, Q.W.; Zhou, X.X.; Zhang, R.Y.; Chen, S.L.; Liu, Q.; Wang, J.; Zhang, Y.; Lin, J.; Xu, J.R.; Gao, Y.J.; et al. Comparison of malignancy-prediction efficiency between contrast and non-contract CT-based radiomics features in gastrointestinal stromal tumors: A multicenter study. *Clin Transl Med* 2020, *10*, e291, doi:10.1002/ctm2.91.
22. Chen, T.; Liu, S.; Li, Y.; Feng, X.; Xiong, W.; Zhao, X.; Yang, Y.; Zhang, C.; Hu, Y.; Chen, H.; et al. Developed and validated a prognostic nomogram for recurrence-free survival after complete surgical resection of local primary gastrointestinal stromal tumors based on deep learning. *EBioMedicine* 2019, *39*, 272-279, doi:10.1016/j.ebiom.2018.12.028.
23. Zheng, J.; Xia, Y.; Xu, A.; Weng, X.; Wang, X.; Jiang, H.; Li, Q.; Li, F. Combined model based on enhanced CT texture features in liver metastasis prediction of high-risk gastrointestinal stromal tumors. *Abdom Radiol (NY)* 2022, *47*, 85-93, doi:10.1007/s00261-021-03321-3.
24. Jager, P.L.; Gietema, J.A.; van der Graaf, W.T. Imatinib mesylate for the treatment of gastrointestinal stromal tumours: best monitored with FDG PET. *Nucl Med Commun* 2004, *25*, 433-438, doi:10.1097/00006231-200405000-00002.
25. Beheshti, M.; Li, S.R.; Vali, R.; Schima, W.; Dudczak, R.; Langsteger, W. The Potential Value of F-18 FDG PET in Comparison to CT in Early Prediction of Response to Imatinib (STI571) Therapy in Patients with Gastrointestinal Stromal Tumors. *IRANIAN JOURNAL OF NUCLEAR MEDICINE* 2007, *15*, 34-42.
26. Prior, J.O.; Montemurro, M.; Orcurto, M.V.; Michielin, O.; Luthi, F.; Benhattar, J.; Guillou, L.; Elsig, V.; Stupp, R.; Delaloye, A.B.; et al. Early prediction of response to sunitinib after imatinib failure by 18F-fluorodeoxyglucose positron emission tomography in patients with gastrointestinal stromal tumor. *J Clin Oncol* 2009, *27*, 439-445, doi:10.1200/jco.2008.17.2742.
27. Miyake, K.K.; Nakamoto, Y.; Mikami, Y.; Tanaka, S.; Higashi, T.; Tadamura, E.; Saga, T.; Minami, S.; Togashi, K. The predictive value of preoperative (18)F-fluorodeoxyglucose PET for postoperative recurrence in patients with localized primary gastrointestinal stromal tumour. *Eur Radiol* 2016, *26*, 4664-4674, doi:10.1007/s00330-016-4242-5.
28. Tirumani, S.H.; Shinagare, A.B.; Jagannathan, J.P.; Krajewski, K.M.; Ramaiya, N.H.; Raut, C.P. Radiologic assessment of earliest, best, and plateau response of gastrointestinal stromal tumors to neoadjuvant imatinib prior to successful surgical resection. *Eur J Surg Oncol* 2014, *40*, 420-428, doi:10.1016/j.ejso.2013.10.021.

Supplementary Materials

Image acquisition parameters			
Portal venous CE-CT	Total (n = 40)	Good respondent (n = 20)	Poor respondent (n = 20)
Manufacturer			
Canon Medical systems	16 (40.0%)	10 (50.0%)	6 (30.0%)
GE Healthcare	8 (20.0%)	3 (15.0%)	5 (25.0%)
Philips Healthcare	3 (7.5%)	0 (0.0%)	3 (15.0%)
Siemens Healthineers	13 (32.5%)	7 (35.0%)	6 (30.0%)
Median peak kilovoltage (kVp)	120.0 (range 100.0 - 140.0)	120.0 (range 100.0 - 140.0)	120.0 (range 100.0 - 135.0)
Median slice thickness (mm)	1.0 (range 0.5 - 5.0)	1.0 (range 0.5 - 5.0)	1.1 (range 0.6 - 5.0)
Median pixel spacing (mm)	0.8 (range 0.6 - 0.9)	0.8 (range 0.6 - 0.9)	0.8 (range 0.7 - 0.9)
[^{18}F]FDG-PET/CT	Total (n = 18)	Total (n = 10)	Total (n = 8)
Manufacturer			
GE Healthcare	1 (5.6%)	1 (10.0%)	0 (0.0%)
Philips Healthcare	13 (72.2%)	8 (80.0%)	5 (62.5%)
Siemens Healthineers	4 (22.2%)	1 (10.0%)	3 (37.5%)
EARL-compliant			
Yes	14 (77.8%)	8 (80.0%)	6 (75.0%)
No	4 (22.2%)	2 (20.0%)	2 (5.0%)
Median activity at start (MBq)	195.6 (range 125.3 - 380.9)	194.1 (range 125.3 - 380.9)	227.8 (range 160.0 - 324.1)
Median PET acquisition (min)	57.0 (range 50.0 - 69.0)	57.0 (range 50.0 - 62.0)	55.5 (range 55.0 - 69.0)
Median slice thickness (mm)	4.0 (range 2.8 - 5.0)	4.0 (range 2.8 - 4.0)	4.0 (range 3.0 - 5.0)
Median pixel spacing (mm)	4.0 (range 2.0 - 4.1)	4.0 (range 2.7 - 4.1)	4.0 (range 2.0 - 4.1)

Table S1. Acquisition parameters for baseline CE-CT and [^{18}F]FDG-PET/CT imaging performed to diagnose the included non-metastatic GIST patients. The values are stratified by responder status, which was determined through volumetric response criteria. Numerical data is presented as median value, along with the concurrent with minimum and maximum range. (CE-CT = contrast-enhanced computed tomography, [^{18}F]FDG-PET/CT = 2-deoxy-2-[^{18}F]fluoro-D-glucose).

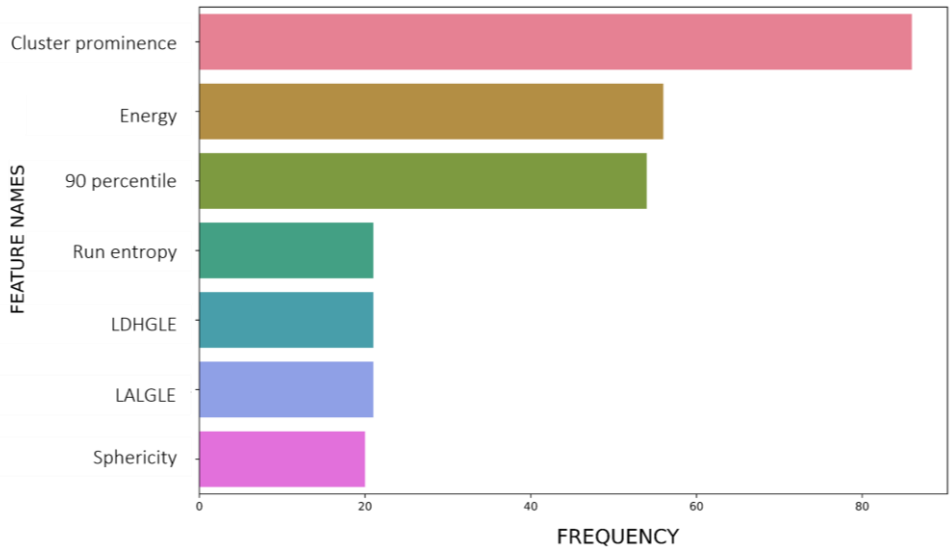

Fig S1. Bar plot visualising the feature selection frequency during the training process involving dataset A (CE-CT imaging features) in a stratified five-fold cross validation and twenty iterations. The three most commonly selected features included cluster prominence, energy and 90 percentile.

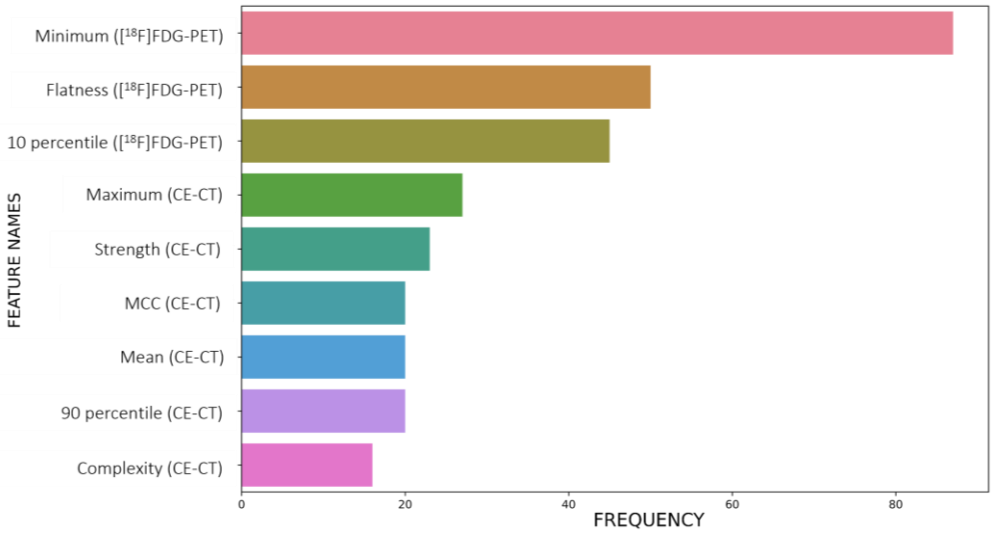

Fig S2. Bar plot visualising the feature selection frequency during the training process involving dataset B (CE-CT and [^{18}F]FDG-PET imaging features) in a stratified five-fold cross validation and twenty iterations. The three most commonly selected features included minimum, flatness and 10 percentile, extracted from [^{18}F]FDG-PET imaging.

Training scores								
	Dataset A (CE-CT)				Dataset B (CE-CT + [^{18}F]FDG-PET)			
Classifiers	Baseline	HP	LASSO	LASSO + HP	Baseline	HP	LASSO	LASSO + HP
LR	0.44 (0.09)	0.43 (0.09)	0.58 (0.10)	0.58 (0.09)	0.30 (0.14)	0.31 (0.14)	0.77 (0.12)	0.76 (0.11)
KNN	0.40 (0.11)	0.47 (0.10)	0.50 (0.13)	0.60 (0.09)	0.35 (0.15)	0.48 (0.12)	0.74 (0.10)	0.81 (0.11)
RF	0.38 (0.11)	0.50 (0.08)	0.51 (0.09)	0.60 (0.09)	0.39 (0.14)	0.67 (0.07)	0.65 (0.14)	0.80 (0.08)
SVM	0.42 (0.12)	0.52 (0.09)	0.58 (0.12)	0.63 (0.09)	0.49 (0.15)	0.68 (0.07)	0.74 (0.11)	0.85 (0.07)

Table S2 F1-scores for four machine learning classifiers during the training process, using dataset A (portal venous CE-CT imaging features) and dataset B (combination of CE-CT and [^{18}F]FDG-PET imaging features) to predict volumetric response in non-metastatic GIST patients. Hyperparameter tuning and feature selection methods (LASSO) were introduced to improve the F1-scores of the classifiers. No performance difference between baseline and hyperparameter tuning was observed for the LR classifier, as the default parameters were used throughout the entire training process. HP = hyperparameter, LR = logistic regression, KNN = k-nearest neighbour, RF = random forest, SVM = support vector machine, AUC = area under the curve.

General discussion

The aim of this thesis was to investigate the additional value of CE-CT and [^{18}F]FDG PET/CT imaging features for the purpose of predicting and monitoring neoadjuvant treatment response in GIST patients. For these patients, monitoring of treatment response is important to eliminate surgical delay, overtreatment, side effects and unnecessary healthcare costs.

1. Definition of response in GISTs

The most complex subject of this thesis remained the definition of 'response'. Response in GISTs can only be determined through radiological criteria, as there is no definitive correlation between pathological response and clinical implications. The percentage of remaining viable tumour cells is not correlated with patient outcome [1]. Although radiological response criteria have been used to predict and monitor response in metastatic GIST patients, the role of radiological response assessments in patients receiving neoadjuvant treatment has yet to be determined. In metastatic patients, small reductions in tumour size (stable disease) are considered favourable, as no disease progression is observed. However, in a neoadjuvant setting, the purpose is to attain significant size reductions of the primary tumour to improve chances of complete surgical resection. As a consequence, larger reductions in tumour size (partial response) are considered beneficial. Nonetheless, TKI treatment can cause tumour destruction without inducing substantial tumour shrinkage and size-based criteria may therefore underestimate the therapeutic effect in GISTs [2,3]. For this reason, we introduced 'surgical benefit' as a new outcome measure for determining response to neoadjuvant treatment. Surgical benefit in this setting is defined by TKI-therapy induced tumour size reductions that ensure tissue and organ preservation and/or improve surgical oversight and planning (i.e., smaller incisions, visualisation of tumour attachment, neutralisation of tumour adhesions to adjacent structures). To better quantify surgical benefit and overcome inter- and intra-observer variability, the authors propose to involve expert opinions of at least two surgical oncologists in order to reach broad consensus.

2. CE-CT imaging

2.1 The RECIST 1.1 and volumetric criteria

When considering CE-CT imaging, RECIST 1.1 and tumour volumetry best predicted surgical benefit in non-metastatic GIST patients treated with neoadjuvant intent. Chen et al. also found a correlation between RECIST 1.1 measurements and a decreased scope of surgery in high risk soft tissue sarcomas [4]. Although this study showed promising results for tumour volumetry to predict surgical outcome, manual segmentation on the images limits its clinical applicability. Namely, before applying volumetric response assessment in clinical practice, a (semi-) automatic segmentation algorithm is preferred to avoid additional labour for radiologists and improve repeatability. RECIST 1.1 measurements performed on the first interim CE-CT scan could accurately predict volumetric response in the non-metastatic GIST patient population. This suggests that RECIST 1.1 measurements can be used to predict effectiveness of TKI therapy in GIST patients at an early phase. Although less accurate than tumour volumetry, RECIST 1.1 could predict surgical benefit with reasonable accuracy. The advantage of RECIST is that radiologists are already used to these criteria in standard clinical practice and it does not lead to extra patient burden. It is therefore recommended to include these measurements in future radiological reports and take these findings into account during multidisciplinary meetings.

2.2 The Choi criteria

In this thesis we conclude that the clinical efficacy of the Choi criteria in determining response in GIST patients is questionable. On theoretical grounds, changes in tumour vascularity induced by TKI treatment

can be measured by reductions in tissue densities, as reflected by CE-CT imaging. However, when looking at the literature, intralesional hypodensity and concurrent heterogeneous enhancement patterns were also observed in high-risk GISTs. A reduction in tumour density could therefore reflect response, as well as aggressive tumour behaviour during longer follow-up periods. These two effects cannot be distinguished on CE-CT imaging and makes clinical decision-making based on these criteria difficult. In addition, technical limitations such as inter-scanner variability, make side-to-side comparisons on tumour density unreliable and impractical.

3. [^{18}F]FDG-PET imaging

Based on the literature review, metabolic changes (standardised uptake value) on [^{18}F]FDG-PET imaging were strongly correlated with tumour response and could aid towards early response prediction and monitoring [5-8]. These metabolic changes preceded morphological changes in size and [^{18}F]FDG-PET imaging may be a useful complement to the routinely used CE-CT imaging modality. In addition to conventional [^{18}F]FDG-PET imaging features, the added value of other features describing patterns and shape of [^{18}F]FDG uptake (radiomic features) was investigated. In our radiomic study, the minimum value on [^{18}F]FDG-PET imaging was significantly higher in the responding group, which can be explained by underlying tumour biology, as more uniform uptake is often associated with less aggressive tumour behaviour [9]. Results should be interpreted carefully as a limited amount of data was included. The clinical added value of [^{18}F]FDG-PET imaging in predicting response to neoadjuvant TKI treatment at baseline by means of radiomics remains unclear.

4. Radiomics

The radiomic models presented in this thesis all scored AUC values below 0.50. However, this does not rule out its potential for early image-based response prediction. Lack of generalisation and concurrent misclassifications may be caused by the limited number of samples and the presence of heterogeneity in the imaging acquisition protocols. In the near future, additional imaging data will be provided to improve the performance of these radiomic models. In addition, to reduce heterogeneity within the imaging acquisition protocols, ComBat harmonisation is warranted. To the author's knowledge, this would be the first CE-CT and [^{18}F]FDG-PET radiomic study predicting volumetric response in GIST patients treated with neoadjuvant intent. Nonetheless, current radiomic studies often assign the same weights for all extracted features and do not correlate their findings with underlying tumour biology. Future research should carry out a more hypothesis-driven radiomic approach, where we select predefined quantitative features from different imaging modalities to reflect intra- and intertumoral heterogeneity.

References

1. Bauer, S.; Hartmann, J.T.; de Wit, M.; Lang, H.; Grabellus, F.; Antoch, G.; Niebel, W.; Erhard, J.; Ebeling, P.; Zeth, M.; et al. Resection of residual disease in patients with metastatic gastrointestinal stromal tumors responding to treatment with imatinib. *Int J Cancer* 2005, *117*, 316-325, doi:10.1002/ijc.21164.
2. Choi, H.; Charnsangavej, C.; de Castro Faria, S.; Tamm, E.P.; Benjamin, R.S.; Johnson, M.M.; Macapinlac, H.A.; Podoloff, D.A. CT evaluation of the response of gastrointestinal stromal tumors after imatinib mesylate treatment: a quantitative analysis correlated with FDG PET findings. *AJR Am J Roentgenol* 2004, *183*, 1619-1628, doi:10.2214/ajr.183.6.01831619.
3. Cohen, N.A.; Kim, T.S.; DeMatteo, R.P. Principles of Kinase Inhibitor Therapy for Solid Tumors. *Ann Surg* 2017, *265*, 311-319, doi:10.1097/sla.0000000000001740.
4. Chen, Y.; Yang, Y.; Wang, C.; Shi, Y. Radiographic response to neoadjuvant therapy and its impact on scope of surgery and prognosis in stage IIB/III soft tissue sarcomas. *BMC Cancer* 2013, *13*, 591, doi:10.1186/1471-2407-13-591.
5. Albano, D.; Bosio, G.; Tomasini, D.; Bonù, M.; Giubbini, R.; Bertagna, F. Metabolic behavior and prognostic role of pretreatment 18F-FDG PET/CT in gist. *Asia Pac J Clin Oncol* 2020, *16*, e207-e215, doi:10.1111/ajco.13366.
6. Hwang, S.H.; Jung, M.; Jeong, Y.H.; Jo, K.; Kim, S.; Wang, J.; Cho, A. Prognostic value of metabolic tumor volume and total lesion glycolysis on preoperative (18)F-FDG PET/CT in patients with localized primary gastrointestinal stromal tumors. *Cancer Metab* 2021, *9*, 8, doi:10.1186/s40170-021-00244-x.
7. Jager, P.L.; Gietema, J.A.; van der Graaf, W.T. Imatinib mesylate for the treatment of gastrointestinal stromal tumours: best monitored with FDG PET. *Nucl Med Commun* 2004, *25*, 433-438, doi:10.1097/00006231-200405000-00002.
8. Prior, J.O.; Montemurro, M.; Orcurto, M.V.; Michielin, O.; Luthi, F.; Benhattar, J.; Guillou, L.; Elsig, V.; Stupp, R.; Delaloye, A.B.; et al. Early prediction of response to sunitinib after imatinib failure by 18F-fluorodeoxyglucose positron emission tomography in patients with gastrointestinal stromal tumor. *J Clin Oncol* 2009, *27*, 439-445, doi:10.1200/jco.2008.17.2742.
9. Miyake, K.K.; Nakamoto, Y.; Mikami, Y.; Tanaka, S.; Higashi, T.; Tadamura, E.; Saga, T.; Minami, S.; Togashi, K. The predictive value of preoperative (18)F-fluorodeoxyglucose PET for postoperative recurrence in patients with localized primary gastrointestinal stromal tumour. *Eur Radiol* 2016, *26*, 4664-4674, doi:10.1007/s00330-016-4242-5.