MSC BIOMEDICAL ENGINEERING

PROGNOSTIC VALUE OF [^{18}F]FDG-PET RADIOMICS TO DETECT PERITONEAL AND DISTANT METASTASES IN LOCALLY ADVANCED GASTRIC CANCER - A SIDE-STUDY OF THE PROSPECTIVE MULTICENTER PLASTIC TRIAL

LIEKE (C.E.) PULLEN

S1798103

16TH NOVEMBER 2022

EXAMINATION COMMITTEE

Chair: Prof dr. L.F. de Geus-Oei
Professor in Radiology focused on Nuclear Medicine at Leiden University Medical Center, professor at Biomedical Photonic Imaging (BMPI) at the University of Twente, l.oei@utwente.nl

Daily supervisor LUMC: W.A. Noortman MSc
Technical Medicine-graduate and PhD student, Biomedical Photonic Imaging (BMPI) at the University of Twente, w.a.noortman@lumc.nl

UT supervisor: Prof. dr. R.H.J.A. Slart,
Biomedical Photonic Imaging (BMPI) at the University of Twente, r.h.j.a.slart@utwente.nl

External supervisor: Dr. M. Dantuma
Multi-Modality Medical Imaging Group (M3I) at the University of Twente, m.dantuma@utwente.nl

LUMC UMC Utrecht

UNIVERSITY OF TWENTE.

LIST OF ABBREVIATIONS

AUC	Area Under The Receiver Operating Characteristic Curve
CT	Computer Tomography
DT	Decision Tree
EANM	European Association of Nuclear Medicine
[^{18}F]FDG	2-[^{18}F]fluoro-2-deoxy-D-glucose
LASSO	Least Absolute Shrinkage and Selection Operator
LR	Logistic Regression
mRMR	maximum Relevance-Minimum Redundancy
PCA	Principal Component Analysis
PET	Positron Emission Tomography
ROC	Receiver Operating Characteristic
RF	Random Forest
SL	Staging Laparoscopy
SUV	Standardized Uptake Value
SVM	Support Vector Machine
VOI	Volume Of Interest

PREFACE

Dear reader,

In this master's thesis, I present the research I have done during my graduation assignment at the section of nuclear medicine at the department of radiology at the University Medical Center Utrecht, in collaboration with the Leiden at the University Medical Center. Over the past half year, I have researched the added value of [^{18}F]FDG-PET-derived radiomics in predicting peritoneal and distant metastases in patients with advanced gastric cancer, as a side-study of the PLASTIC trial. Radiomics is a field of research that aims to extract quantitative metrics of imaging data that are difficult to recognize by the human eye. It is believed that it can build prediction models to guide clinical decision-making.
The thesis starts with a general introduction explaining the basics of gastric cancer, PET and radiomics. An article follows, presenting the methodology, results and a part of the discussion. The discussion is continued in the general discussion, giving more information about the methodology and future perspectives on radiomics. Lastly, I have written a small acknowledgement.

I hope you will enjoy reading this thesis.

Lieke Pullen, December 2022

GENERAL INTRODUCTION

Gastric cancer

Epidemiology of gastric cancer

In 2020, the incidences of gastric adenocarcinoma and cardia cancer in the Dutch population were over 1100 and 512, respectively[1]. It is the 6th most common tumour type, and the 3th most known reason for cancer-related death[2]. Assuming current incidence rates and population growth of 2020 remain the same, the annual number of new gastric cancer cases is predicted to increase by 62% from 1.09 million cases to 1.77 million cases by 2040[3]. In 2040, the expected global burden will be around 28.4 million cases of gastric cancer[2].

Differences between the Asian and Western populations are long known. Globally, the Asian population has a much higher gastric cancer incidence, and mortality compared to the Western population[4]. Risk factors such as smoking, alcohol intake and food preservation increase the chance of developing gastric cancer. Especially the latter is a high-risk factor in Asian countries where food is traditionally preserved with salt. Screening programs are currently mainly used in Japan and South Korea[5] due to their high incidence, and resulted in a 40% reduction in gastric-cancer mortality[6]. Since a screening program is not implemented in the Dutch system[1], gastric cancer in the Dutch population is discovered at a later stage in comparison to the Asian population.

Diagnosis and staging

Gastric cancer has multiple common symptoms, including indigestion, poor appetite, weight loss, melaena and abdominal pain. Generally, a primary gastric tumour is diagnosed by endoscopic examination, where the tumour is located within the stomach, and a histological biopsy is taken to confirm the presence of gastric adenocarcinoma[7]. Consequently, a computer tomography (CT) scan of the abdomen and chest is performed to stage the disease. Since screening is not routine in the Netherlands, symptoms of gastric cancer usually occur at a later stage.

Dependent on the penetration of the tumour within the stomach walls and whether or not metastases are present, the tumour is assigned to a particular stage. The anatomy of the stomach wall is given in figure 1.
Staging of gastric cancer is defined by the depth of tumour invasion of the primary tumour (T), presence of locoregional lymph node metastases (N) and presence of any distant and/or peritoneal metastases (M). The specific descriptions of T-, N- and M-stages are given in table 1. The classification of staging defined by the Union for International Cancer Control is given in table 2[10]. When distant metastases are present, gastric cancer is defined as stage IV and considered inoperable for radical intent.

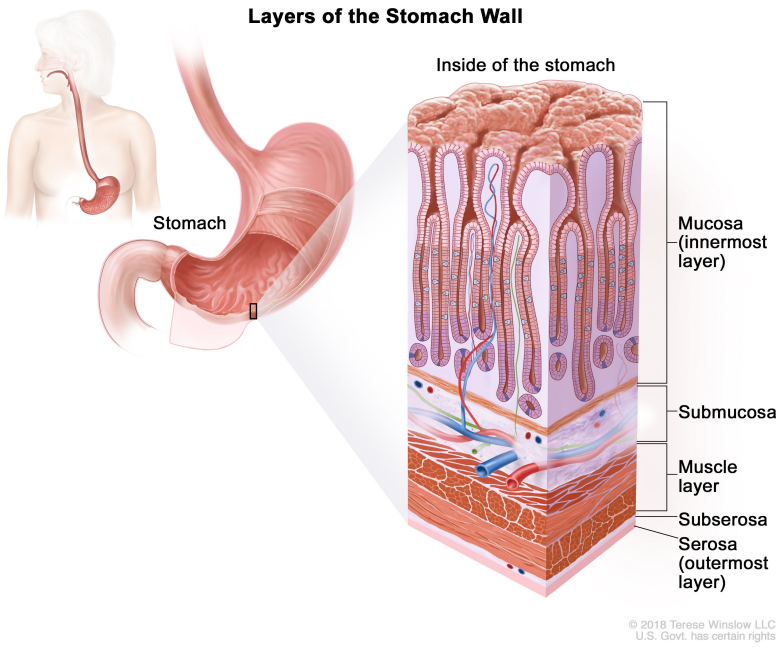


Figure 1: Layers of the stomach wall. The wall of the stomach is made up of the mucosa (innermost layer), submucosa, muscle layer, subserosa, and serosa (outermost layer)[9].

T-stage	Description
Tis	Intraepithelial tumour without invasion of the lamina propria, high-grade dysplasia
T1a	Tumor invades lamina propria, muscularis mucosae, or submucosa.
T1b	Tumor invades submucosa.
T2	Tumor invades muscularis propria.
T3	Tumor penetrates the subserosal connective tissue without invasion of the visceral peritoneum or adjacent structures.
T4a	Tumor invades serosa (visceral peritoneum)
T4b	Tumor invades adjacent structures/organs

N-stage	Description
N0	No regional lymph node metastasis.
N1	Metastases in 1 or 2 regional lymph nodes.
N2	Metastases in 3 to 6 regional lymph nodes.
N3a	Metastases in 7 to 15 regional lymph nodes.
N3b	Metastases in 16 or more regional lymph nodes.

M-stage	Description
M0	No distant metastases.
M1	Distant metastases.

Table 1: Classification of gastric tumors: T = primary tumor invasion depth, N = regional lymph node, M = distant metastases;[10]

	N0	N1	N2	N3a	N3b	M1
T1a	IA	IB	IIA	IIB	IIIB	IV
T1b	IA	IB	IIA	IIB	IIIA	IV
T2	IB	IIA	IIB	IIIA	IIIB	IV
T3	IIA	IIB	IIIA	IIIB	IIIC	IV
T4a	IIB	IIIA	IIIB	IIIC	IIIC	IV
T4b	IIIA	IIIB	IIIB	IIIC	IIIC	IV
M1	IV	IV	IV	IV	IV	IV

Table 2: Classification of gastric cancer stage, defined by the T-, N- and M-stage of the tumour. If there are any distant metastasis, the cancer is defined as stage IV.[10]

Gastric cancer can be classified into two main types according to Lauren classification: intestinal and diffuse-type carcinoma. Intestinal-type tumours exhibit adhesion and are arranged in tubular or glandular formations. Diffuse-type tumours lack adhesion, leading to the gathering of non-cohesive tumour cells[11]. The tumours differ in pathology, epidemiology, and aetiology, and their classification might be the basis of individual chemotherapy for gastric cancer[12].

Metastases of gastric cancer can be divided into two categories: peritoneal and distant metastases. Peritoneal metastases refer to metastatic involvement of the peritoneum[13]. In 14% of the advanced stage (stage IV) gastric cancer, peritoneal metastases are already present at first presentation[14]. Distant metastases are metastases in other organs by hematogenous dissemination or distant lymph nodes via lymphomatous dissemination. Distant metastases of gastric cancer are most often found in the liver[15].

Treatment

Treatment of gastric cancer is dependent on the tumour stage. In stage I-III gastric cancer, surgical resection - gastrectomy - of the tumour is often the preferred method[16]. In early-stage patients with intramucosal tumours, surgical resection can be performed endoscopically.
Surgery can be combined with perioperative or adjuvant chemotherapy, or chemoradiotherapy[17]. In stage IV, where the tumour is considered inoperable, palliative care is provided because the disease is no longer curable. This care can consist of palliative systemic therapy, palliative immunotherapy, combinations of these two, or best supportive care. Palliative therapy focuses on symptom management, pain control, and patient comfort and may be done to prolong life.

Positron Emission Tomography

Positron emission tomography (PET) is a major functional imaging technique in nuclear medicine[19]. PET is used to image the distribution of a radionuclide administered in the body.
The imaging technique is based on the coincidence of two photons. In PET imaging, a radioactive isotope is intravenously injected into the patient[18]. The isotope undergoes decay leading to a positron being emitted, which has the opposite charge as an electron. After a short travel through the tissue, the positron loses most of its energy and combines with an electron. This process is called annihilation, converting the particle's mass to radiant energy. It produces two annihilations (gamma) photons of 511 keV moving in the opposite direction, which the scanner can detect. This event is called annihilation coincidence detection.

The crystals within the PET detector absorb the photons, producing light that is converted to an electrical signal (figure 2). PET counts the number of detected photons. Since the detected

photons are annihilated in the opposite direction, the line including both photons can be defined by a line of response (LOR). With image reconstruction, the position of the coincidence events can be determined.

The raw data from the PET scan are stored in a list mode file, which next to the coordinates of the position for the individual events, stores the millisecond interval at which the event occurred[20]. The image can be reconstructed from the list mode file using iterative reconstruction algorithms. In such algorithms, an initial guess of the radioactivity distribution is made and compared to the measured data. From this comparison, an image correction factor is determined and applied to the reconstructed image. This new guess is then compared to the measured data, and so on, until the reconstructed image is equal to the measured image.

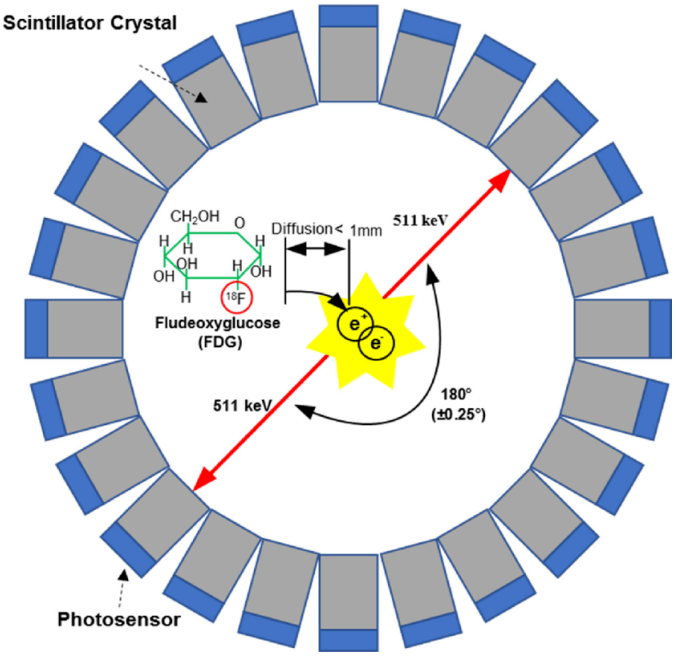


Figure 2: The basic principle of a positron emission tomography (PET) system: A PET detector ring detects a pair of gamma photons with an energy of 511 keV (red arrows) which results from the annihilation of an electron with a positron emitted by the radiotracer ([$_{18}$F]FDG).[19]

[^{18}F]FDG

2-[^{18}F]fluoro-2-deoxy-D-glucose ([^{18}F]FDG) is a radiotracer used for PET imaging and is a glucose analogue, wherein the glucose molecule, the -OH group is replaced with a radioactive ^{18}F. Due to its relatively short half-life of 109.7 minutes (approximately two hours), it has a relatively low radiation dose but enough dose for imaging. The injected activity is determined using a dosage regimen based on the patient's body weight. Since its attachment to glucose, the radiotracer is taken up by living cells via glucose transporters on the cell membrane using hexokinase. [^{18}F]FDG uptake reflects the tissue glucose metabolism.

Cancer cells generally take up more glucose than healthy cells, and [$_{18}$F]FDG accumulation in cells is proportional to its utilisation. More [$_{18}$F]FDG will thus be present in cancer cells and other organs and tissues with high metabolisms, such as the brain, the liver and the kidneys or other sites with active inflammation. After ^{18}F has decayed, the deoxyglucose molecules will become part of the glycolytic pathway[21].

Computed Tomography

Computed Tomography (CT) is a 3D anatomical imaging technique that reflects X-ray attenuation by different tissues. CT uses quickly rotating x-ray tubes and an oppositely placed row of detectors placed in a gantry to produce an attenuation profile of the patient. It consists of a varying number of detector rows (8-64), which hold between 1000 and 2000 detectors per row. An x-ray source moves around the patient, and with each rotation, a 2D image slice of the patient is constructed. The intensity of the x-ray beam depends on the total attenuation of the tissue it passes. Every slice leads to another attenuation profile stored in a sinogram.

Radiomics

Radiomics is a field of research that aims to extract quantitative metrics of imaging data that are difficult to recognise by the human eye[22]. When combined with clinical, genomic or histologic data, it may give information about tumour characterisation, treatment response prediction or other tumour characteristics. It is believed that it can build prediction models in order to guide clinical decision-making.

The pipeline of radiomics starts with image acquisition and reconstruction, including attenuation correction, artefact correction, and filtering. From these images, a volume of interest (VOI) is segmented either manually or (semi-)automatically. The VOIs are used for feature extraction. The next step is feature selection, where unimportant features are discarded to reduce the dimensions and more important features remain. A model is trained by implementing the selected features to predict an outcome measure. After training based on the selected features, the model is validated to investigate its performance. With a high-performance model, it can predict the outcome for new data sets[23].

Feature extraction

Radiomics pipelines can be divided into two categories: handcrafted and deep learning pipelines (figure 3). Handcrafted radiomics is considered the traditional form. In this pipeline, predefined features are extracted from the VOI. With deep learning radiomics, raw images are used and the extraction, selection, training and classification are performed by the algorithm.[24].

Three types of handcrafted features are extracted from the imaging data: shape features, intensity features, and texture features. Shape features describe the shape and size of the VOI. Intensity features quantify the tracer uptake within the lesion but do not describe any heterogeneity within the VOI. The intensity features can be quantitatively described by the standardised uptake value (SUV) and are commonly used in PET studies[25].
The SUV expresses the regional tracer uptake by the lesion, normalised by the injected radioactivity and weight of the patient. Its variations, such as SUV_{min} (minimum SUV), SUV_{max} (maximum SUV) and SUV_{peak} (average SUV within 1 cm^3 of the VOI[26]) can objectively give these features. Features such as skewness and kurtosis describe the shape of the intensity distribution with the VOI and are also commonly used parameters to describe intensity features[22]. The heterogeneity of the tracer uptake is described by the last category of features: texture features. These features express the relationship between neighbouring voxels and thus say something about the spatial heterogeneity of the tracer uptake, such as Small Area Emphasis. This feature measures the distribution of small size zones, with the size zone being a number of connected voxels. A greater value indicates smaller size zones and more fine textures[27]. A schematic overview of the pipelines and features is given in figure 3.

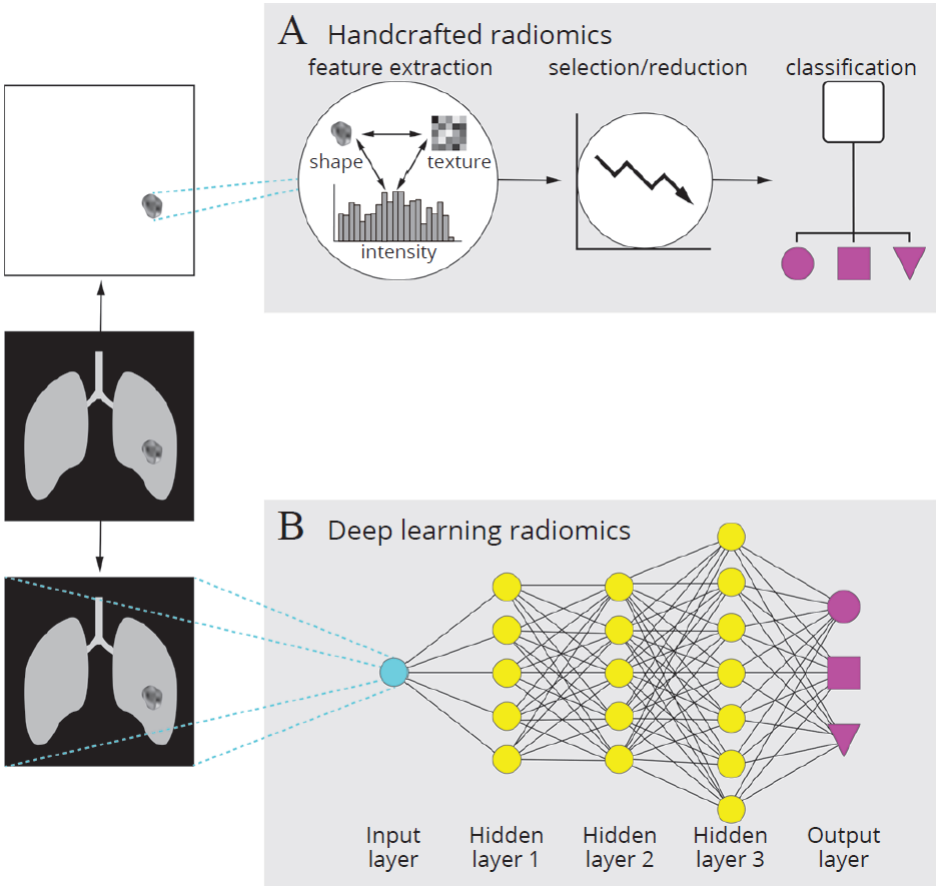


Figure 3: Schematic overview of feature selection in A) handcrafted and B) deep learning pipelines.[24]

Feature selection or dimensionality reduction

Since hundreds or even thousands of features can be extracted from each VOI, not all features hold important information about the outcome measure of interest. An important challenge of radiomics is the large number of features compared to the number of patients, which introduces the curse of dimensionality[28]. The curse refers to analysing high-dimensional spaces on a small amount of data. As the number of radiomics features increases, the data space increases exponentially. When applied to a small amount of data, this leads to overfitting of the model[29]. Furthermore, the computational time and storage requirements increase when implementing more variables, from which deep learning suffers more than handcrafted radiomics. The balance between the right amount of features regarding the amount of data is fragile, and of great importance for a good performing model[30]. Therefore, reducing the number of features is important to prevent the overfitting of the radiomics models. Overfitting occurs with a large number of features for a relatively small number of patients and may lead to unbalanced results[31]. Feature selection can be performed supervised by the outcome measure or unsupervised.

In supervised feature selection, labelled data sets with a known outcome measure are used for feature selection. The outcome measure indicates whether or not the lesion belongs to a certain category. The algorithm then tries to find features that predict the outcome measure of lesions[23].
Supervised methods include least absolute shrinkage and selection operator (LASSO) and maximum relevance-minimum redundancy (mRMR). LASSO uses regression analysis on the train-

ing data and ranks the features based on their predictive accuracy via parameter λ. After regression, most of these coefficients will shrink to zero, and the remaining are selected as predicting features[32]. mRMR uses a similar method, adding a feature to a set of predictive features. However, if the feature contradicts another feature in the set, it will not be added[31].

With unsupervised feature selection or dimensionality reduction, the outcome is unknown. The algorithm will process unlabeled data and structure it into clusters or relationship patterns. The new data will be assigned to clusters rather than categories.
Unsupervised methods for feature selection include principal component analysis (PCA) and clustering correlation matrix. PCA aims to describe the total variation in the data set from only a few latent components and uses these to indicate grouping within the data set. Clustering clusters the data into groups with high intertwined correlation but with a little or high correlation to the other clusters (depending on the type of clustering). It is mostly visualised in a correlation map[33].

Classification algorithms

After feature extraction and selection, classification of the features is necessary to train the model correctly[34]. The data set is split into a training and validation set. The training set is used for training the model, and the validation set is used to classify the model's performance.

- Logistic regression (LR): finds the probability that data belongs to a certain category and gives it a score between 0 and 1. A threshold must be installed to assign the data to the different categories.
- Support vector machine (SVM): the number of features are mapped onto an equal number of dimensions. Then, it divides the features into two classes with maximum separation between the two.
- Decision tree (DT): classifies data into a tree-like structure with multiple levels. Each internal node represents a 'choice', redirecting to the sub-level with another node etc.
- Random forest (RF): very similar to DT but less sensitive to changes in training data. It divided the data into several DTs, with each tree using different parts of the training data set. The forest chooses the classification by majority votes of all DTs.

PLASTIC study

In the Netherlands, 60% of the patients[35] with gastric cancer undergo a curative gastrectomy since most tumours are irresectable or metastases are present at first clinical presentation. In addition, the prognosis after curative treatment is poor, with a five-year survival rate of 36-45%[36][37]. As mentioned earlier, gastrectomy is the preferred method but might not be the best treatment for the patients considering the relatively poor survival rate is mainly due to the recurrence of the disease.
A gastrectomy can be considered futile if unexpected intraoperative peritoneal metastases or an irresectable tumour are detected, or when recurrence or distant metastases occur shortly after gastrectomy (within six months). Such situations lead to inappropriate health care because patients undergo curative treatment with substantial morbidity, mortality and costs related to gastrectomy since gastric cancer surgery cannot cure their disease. Hence, accurate preoperative evaluation of tumour resectability, the presence of metastases and the risk of recurrence is of great importance to reduce the number of futile gastrectomies and improve individual patient care in gastric cancer. This is currently done with a gastroscopy to determine the size and location of the tumour and take a biopsy, and CT, to detect metastases.

Preoperative evaluation of tumour resectability, the presence of metastases and the risk of recurrence is therefore of great importance to reduce the number of futile gastrectomies. This is currently done with a gastroscopy[17] to determine the size and location of the tumour and take a biopsy, and CT, to detect metastases. Unfortunately, as mentioned, the sensitivity of CT is low[38]. Currently, CT of the thorax and abdomen is performed to detect metastases. However, this modality's sensitivity for peritoneal metastasis is low (0.33, 95% CI 0.16-0.56)[39].

Because CT has suboptimal accuracy in detecting distant and peritoneal metastases[8], according to international guidelines, PET scans and staging laparoscopy (SL) were implemented in Dutch hospitals[17]. The aim is to stage locally advanced gastric cancer (T3-4 and/or cN+) to detect occult metastases that were not diagnosed with CT[7]. Previous studies showed that [^{18}F]FDG-PET/CT improves the sensitivity of detecting distant metastases but still has poor sensitivity to detect peritoneal metastases (0.28, 95% CI 0.17-0.44)[39]. Staging laparoscopy can detect peritoneal metastases and/or irresectable gastric cancer. After detecting distant/peritoneal metastases or irresectable disease, the treatment strategy is changed from curative treatment to treatment with palliative intent.

The PLASTIC study by Gertsen et al. was conducted between August 2017 and February 2020 and included 394 patients[40]. This research evaluated the added value of [^{18}F]FDG-PET/CT and SL, in addition to initial staging through gastroscopy and CT in patients with locally advanced gastric cancer. [^{18}F]FDG-PET/CT detected distant metastases in 3% of the patients. In contrast, SL detected peritoneal metastases and/or locally irresectable disease in 19% of the patients, with a 2% overlap between [^{18}F]FDG-PET/CT and SL where both modalities detected non-curable disease. In 16% of the patients, the treatment intent was changed from curative to palliative care. This study concluded that [^{18}F]FDG-PET/CT has limited added value, whereas SL considerably reduced futile gastrectomies due to detecting non-curable disease. Therefore, routine use of [^{18}F]FDG-PET/CT is not recommended, whereas SL should be routinely performed in staging locally advanced gastric cancer.

cancers

MDPI

Article

Prognostic value of [18F]FDG PET radiomics to detect peritoneal and distant metastases in locally advanced gastric cancer - a side-study of the prospective multicenter PLASTIC trial

Lieke C.E. Pullen[1,*], Wyanne A. Noortman[1,2], Lianne Triemstra[3], Cas de Jongh[3], Fenna J. Rademaker[4], Romy Spijkerman[4], Gijsbert M. Kalisvaart[1], Emma C. Gertsen[3], Lioe-Fee de Geus-Oei[1,2], Nelleke Tolboom[5], Wobbe O. de Steur[6], Maura Dantuma[7], Riemer H.J.A. Slart[2,8], Richard van Hillegersberg[3], Peter D. Siersema[9], Jelle P. Ruurda[3], Floris H.P. van Velden[1,†], Erik Vegt[10,†], on behalf of the PLASTIC Study Group[‡]

[1] Department of Nuclear Medicine, Department of Radiology, Leiden University Medical Center, Leiden, The Netherlands; w.a.noortman@lumc.nl (WAN); f.h.p.van_velden@lumc.nl (FHPvV); g.m.kalisvaart@lumc.nl (GMK); L.F.de_Geus-Oei@lumc.nl (LFdGO);
[2] Department of Biomedical Photonic Imaging Group, University of Twente, Enschede, the Netherlands.
[3] Department of Surgery, University Medical Center (UMC) Utrecht, Utrecht, The Netherlands. L.Triemstra@umcutrecht.nl (LT); C.deJongh@umcutrecht.nl (CdJ); E.C.Gertsen-2@umcutrecht.nl (ECG); R.vanHillegersberg@umcutrecht.nl (RvH); J.P.Ruurda@umcutrecht.nl (JPR);
[4] TechMed Centre, University of Twente, Enschede, The Netherlands. f.j.rademaker@student.utwente.nl (FJR); r.spijkerman@student.utwente.nl (RS)
[5] Department of Nuclear Medicine, Department of Radiology, University Medical Center (UMC) Utrecht, Utrecht, The Netherlands. N.Tolboom@umcutrecht.nl (NT);
[6] Department of Surgery, Leiden University Medical Center, Leiden, the Netherlands. w.o.de_steur@lumc.nl (WdS);
[7] Multi-Modality Medical Imaging Group, TechMed Centre, University of Twente, Enschede, The Netherlands. m.dantuma@utwente.nl (MD);
[8] Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. r.h.j.a.slart@umcg.nl (RHJAS);
[9] Department of Gastroenterology, Radboud University Medical Center, Nijmegen, the Netherlands. peter.siersema@radboudumc.nl (PDS);
[10] Department of Radiology and Nuclear Medicine, Erasmus MC University Medical Center Rotterdam, Rotterdam, the Netherlands. e.vegt@erasmusmc.nl (EV);
* Correspondence: lieke.ce.pullen@gmail.com;
† These authors contributed equally to this work;
‡ Collaborators of the PLASTIC Study Group is provided in the Supplementary Material

Simple Summary: Patients with locally advanced gastric cancer have a five-year survival rate of 36-45% after curatively intended D2-gastrectomy combined with perioperative chemotherapy. This relatively poor survival is mainly due to the recurrence of the disease. This study aimed to improve the detection of distant and peritoneal metastases on [18F]FDG-PET images in patients with advanced gastric cancer by extracting and analysing imaging features from these images, known as radiomics. Three predictive models were developed to determine the added value of implementing radiomics: with clinical variables only, radiomic features only, and a combination of both. [18F]FDG-PET-based radiomics showed no additional value in predicting distant nor peritoneal metastases in locally advanced gastric cancer patients.

Abstract: Distant and peritoneal metastases from gastric cancer are frequently missed on CT and [18F]FDG-PET/CT during clinical and radiological evaluation, resulting in patients undergoing unnecessary D2-gastrectomy. This study aimed to improve the prediction of distant and peritoneal metastases on [18F]FDG-PET by implementing radiomics. Data from 18 Dutch hospitals participating in the prospective multicenter PLASTIC-study were used to develop the radiomics model. From [18F]FDG-PET images of 206 patients, 105 radiomic features were extracted. Three machine-learning models were constructed to predict metastases from the extracted features: 1)

with clinical variables only, 2) with radiomic features only, and 3) a clinicoradiomic model. Oversampling was performed to reduce class imbalance. In addition, subgroup analysis based on Lauren classification was performed. Spearman rank's correlation and least absolute shrinkage and selection operator regression were used for feature selection, and logistic regression was used as a classifying algorithm. Performance was scored based on the area under the curve (AUC) and accuracy. Sham experiments were performed to validate the findings. Of the 206 patients included, 21% had confirmed peritoneal or distant metastases. For the complete dataset, no model correctly predicted metastases (AUC of 0.31, 0.35 and 0.30, respectively). Neither of the models predicted metastases correctly in intestinal/mixed and diffuse-type tumours. None of the models correctly predicted distant and peritoneal metastases from advanced gastric cancer. Overall, [^{18}F]FDG PET-based radiomics showed no added value compared to the clinical model.

Keywords: [^{18}F]FDG-PET/CT; gastric cancer; radiomics; metastases

1. Introduction

Gastric cancer is the 3rd most common cause of cancer-related death worldwide [1]. The prognosis after curatively intended treatment is relatively poor, with a five-year survival rate of 36-45% after perioperative chemotherapy combined with D2-gastrectomy. The main reasons for failure of curative treatment are distant metastases during neoadjuvant chemotherapy, unexpected intraoperative peritoneal metastases or tumour irresectability at the onset of gastrectomy, or local or distant recurrences shortly after surgery [2][3]. In the Netherlands, only 60% of gastric cancer patients undergo curative D2-gastrectomy since the remaining patients present with irresectable tumours or distant metastases [4]. After detecting distant metastases or irresectable disease, treatment is changed from curative to palliative intent. Hence, accurate primary staging is crucial.

Currently, computed tomography (CT) of the thorax and abdomen is performed to detect any metastases. However, the sensitivity of CT for detecting peritoneal carcinomatosis (22-33%) and distant metastases (33%, 95% CI: 16--56%) is low [5]. The Dutch multicenter PLASTIC study assessed the diagnostic performance and clinical and financial impact of 2-[^{18}F]fluoro-2-deoxy-D-glucose positron emission tomography combined with CT ([^{18}F]FDG-PET/CT) and staging laparoscopy (SL), in addition to initial staging with CT for locally advanced gastric cancer (cT3-4 and/or cN+) as the standard of care according to Dutch national guidelines [4][6]. Nevertheless, the sensitivity of visual radiological assessment of [^{18}F]FDG--PET/CT for the detection of distant metastases was only 33% (95% CI: 17-53%). The PLASTIC-study did not find additional value of qualitative assessment of [^{18}F]FDG-PET/CT in gastric cancer. However, medical images might contain more information than can be assessed visually by a human reader (experienced nuclear medicine physician), and a more in-depth quantitative assessment might be of added value [7]. Radiomics, the extraction of large amounts of quantitative imaging features from medical imaging and subsequent modelling to associate the imaging features with clinical outcome measures, is a rapidly evolving field in medical imaging [8][9]. Radiomics aims to find stable and clinically relevant image-derived biomarkers and imaging markers that may provide new insights into tumour biology and guide patient management.

Several previous studies showed promising results of CT-based radiomics for the identification of metastases in gastric cancer [10-12], but only a few investigated the predictive value of [^{18}F]FDG-PET/CT radiomics [13][14]. Hence, the added value of FDG-PET/CT-radiomics for gastric cancer is unclear.

This study aimed to assess the added value of [^{18}F]FDG-PET-based radiomics and clinical characteristics for identifying peritoneal and distant metastases in patients with advanced gastric cancer.

2. Materials and Methods

Patient inclusion

[^{18}F]FDG-PET/CT scans were collected from the PLASTIC study hospitals [15]. Only patients with an [^{18}F]FDG-avid primary tumour were included and readily available for radiomic analysis before October 1, 2022. A non-WMO declaration (METC 16-633/C) had been obtained from the Medical Ethical Review Board of the University Medical Center Utrecht. In addition, the trial was approved by the institutional review boards in each of the 18 participating centres. Written informed consent was obtained from all patients. The patient population consisted of patients with surgically resectable, advanced gastric adenocarcinoma (cT3-4b, N0-3, M0) who were scheduled for treatment with curative intent after initial staging with gastroscopy and CT. Patients first underwent an [^{18}F]FDG-PET/CT, followed by a staging laparoscopy if [^{18}F]FDG-PET/CT was negative, as the standard of care according to Dutch national guidelines [4]. The presence of metastases was confirmed based on (histo)pathological biopsy/cytology and/or follow-up imaging.

Image acquisition & reconstruction

The [^{18}F]FDG-PET/CT acquisitions were performed using site-specific scanning protocols, with most hospitals in the Netherlands following the European Association of Nuclear Medicine (EANM) guidelines version 1.0 for tumour PET imaging [15]. When these EARL-compliant PET images were unavailable, the PET images included were reconstructed according to the site-specific reconstruction protocol. Patients had to refrain from exercise and fast for at least 4 to 6 hours before injecting [^{18}F]FDG. Patients were prehydrated by drinking approximately 1 litre of water in the 2 hours before injection. Fasting blood glucose had to be preferably below 11 mmol/L. After the injection of [^{18}F]FDG, patients remained seated or lying and silent for 1 hour in a warm room. The acquisition of a PET scan from the eyes to thighs started approximately 60 min the injection of [^{18}F]FDG, being accompanied by a low-dose CT of the same scanning range [6].

VOI delineation

Volume of interest (VOI) delineation was performed using 3DSlicer (version 4.11.2, www.slicer.org) with the extensions PET-IndiC and PETDICOMExtension, and in-house built software implemented in Python (version 3.7, Python Software Foundation, Wilmington, Delaware, USA) [16][17]. VOIs were delineated using an isocontour that applies an adaptive threshold of 50% of the peak standardized uptake value (SUV_{peak}), obtained using a sphere of 12 mm diameter [18], corrected for local background [19]. In addition, boxing was applied to exclude surrounding [^{18}F]FDG-avid tissues.

Radiomic pipeline

Radiomic features were extracted from the VOI using PyRadiomics (version 3.0.1) in Python (version 3.7) [20]. For each VOI, the standard feature set of 105 radiomic features was extracted: 18 first order, 14 shape, 24 grey level co-occurrence matrix (GLCM), 16 grey-level run-length matrix (GLRLM), 16 grey-level size zone matrix (GLSZM), 14 grey-level dependence matrix (GLDM) and 5 neighbouring grey tone difference (NGTDM) features (Appendix 1). In addition, a fixed bin width of 0.5 g/mL was applied [21]. Furthermore, the VOIs were interpolated to isotropic voxels of 4x4x4 mm^3 to meet the EANM guidelines, using B-spline interpolation (grids aligned by centre, PyRadiomics default).

Statistical analysis

To predict the presence of metastases from the extracted radiomic features and clinical variables, a clinical model, radiomics model and clinicoradiomics model was built. A Kolmogorov-Smirnov test was performed to test the significance of the SUVmean between the EARL-compliant and non-EARL-compliant scans [22]. ComBat harmonisation was applied between radiomic features derived from EARL-compliant and

non-EARL-compliant scans to reduce variability by scanning protocols using neuroCombat in Python (version 3.9) [23].

The dataset was split into a training and test set (80:20), stratified for the presence of distant metastases. Radiomic features were scaled (μ=0, σ=1) to prevent a large contribution of high-valued features. To exclude features with high mutual correlations, features were first clustered into 40 clusters based on Spearman's rank correlation. The features with the best representation of the individual cluster were used as input for least absolute shrinkage and selection operator (LASSO) regression [24]. Features were further reduced using LASSO by calculating hyperparameter λ using stratified 5-fold cross-validation. As recommended in the literature, selected radiomic features were introduced in a Logistic Regression algorithm to identify metastases [25]. A clinical model was created with the variables: clinical T-stage, clinical N-stage, Lauren classification, degree of differentiation and Her2Neu-status [26]. The clinical variables Lauren classification, degree of differentiation and Her2Neu-status were imputed based on fourteen other variables (Appendix 2) due to a high percentage of unknown classes (21%, 43% and 54%, respectively). The clinicoradiomics model was created with all clinical variables and radiomic features selected based on the absolute regression coefficient. To compensate for the class imbalance for metastases, the training set was oversampled randomly and compared to the analysis with a non-oversampled training set. The function oversamples the minority class by picking samples at random with replacement [27].

Classification performances were presented by the area under the curve (AUC) (AUC of 0.5 represents no discrimination, AUC between 0.5 and 0.7 poor discrimination, AUC between 0.7 and 0.8 acceptable discrimination, AUC between 0.8 and 0.9 excellent discrimination and an AUC of 0.9 and higher outstanding discrimination) and accuracy for both the training set as the validation set.

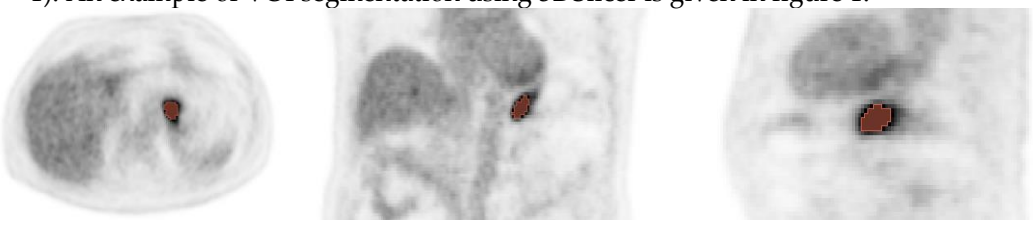
Subgroup analyses were performed based on the Lauren classification [26]. Tumours with mixed classification were considered intestinal tumours based on their similarity in pathology.

A sham experiment was performed to validate the findings and to eliminate spurious correlations and confounding effects in radiomic studies [28]. To that end, the outcome labels were randomly shuffled for 100 iterations. Randomisation of the outcome labels preserves the multicollinearity and distributions of the radiomic features and the prevalence of the outcome, but it uncouples their potential relation. In the sham experiment, an AUC of 0.50 was expected.

3. Results

Patient characteristics

A total of 236 patients were considered for radiomic analysis, of which thirty were excluded. Reasons for exclusion were: doubtful tumour segmentation (n=12), corrupt DICOM files (n=10) or missing clinical variables (n=8). Thus 206 remaining patients with advanced gastric adenocarcinoma were analyzed, of which 43 (21%) had metastases (table 1). An example of VOI segmentation using 3DSlicer is given in figure 1.



(**a**) (**b**) (**c**)

Figure 1. VOI segmentation of the tumors in (a) transversal, (b) coronal and (c) sagittal direction. The VOI derived with an adaptive threshold method is shown in red and is used for radiomic feature extraction.

Table 1. Baseline characteristics of the 206 included patients, divided in a training set (n=165) and a validation set (n=41) Statistically significant differences are indicated in bold.
[a] Mann-Whitney U test
[b] Fisher's exact test
[c] Pearson's chi-squared test

Characteristic (n (%))	**Training set (n=165)**	**Validation set (n=41)**	***p*-value**
Age (years), mean ± standard deviation)	68 ± 10	66 ± 11	0.35[a]
Sex			
Male	101 (61%)	29 (71%)	1[b]
Female	64 (39%)	12 (29%)	
Presence of metastases			
Yes	35 (21%)	8(20%)	1[b]
No	130 (79%)	33 (80%)	
Clinical T-stage			
T3	13 (8%)	2 (5%)	
T4a	121 (73%)	35 (85%)	**0.009**[c]
T4b	29 (18%)	4 (10%)	
Missing	2 (1%)	0 (0%)	
Clinical N-stage			
N0	80 (48%)	16 (39%)	0.12[c]
N+	81 (49%)	25 (61%)	
Missing	4 (2%)	0 (0%)	
Tumour location			
Cardia	30 (18%)	10 (24%)	
Corpus & fundus	46 (28%)	11 (27%)	0.48[c]
Antrum & pylorus	66 (40%)	19 (46%)	
Diffuse	18 (11%)	1 (2%)	
Missing	5 (3%)	0 (0%)	
Lauren classification*,†			
Intestinal	98 (59%)	27 (66%)	0.51[b]
Diffuse	67 (41%)	14 (34%)	
Differentiation†			
Well	9 (5)	0 (0%)	
Moderate	72 (44%)	15 (37%)	0.37[c]
Poor	81 (49%)	26 (63%)	
Undifferentiated	3 (2%)	0 (0%)	
Her2Neu status†			
Positive	12 (7%)	1 (2%)	1[b]
Negative	153 (93%)	40 (98%)	
EARL-compliant PET scan			
Yes	74 (45%)	20 (49%)	0.12[b]
No	91 (55%)	21 (51%)	

* Tumors with mixed classification were considered intestinal tumors based on their similarity in pathology.
† Characteristics were imputed.

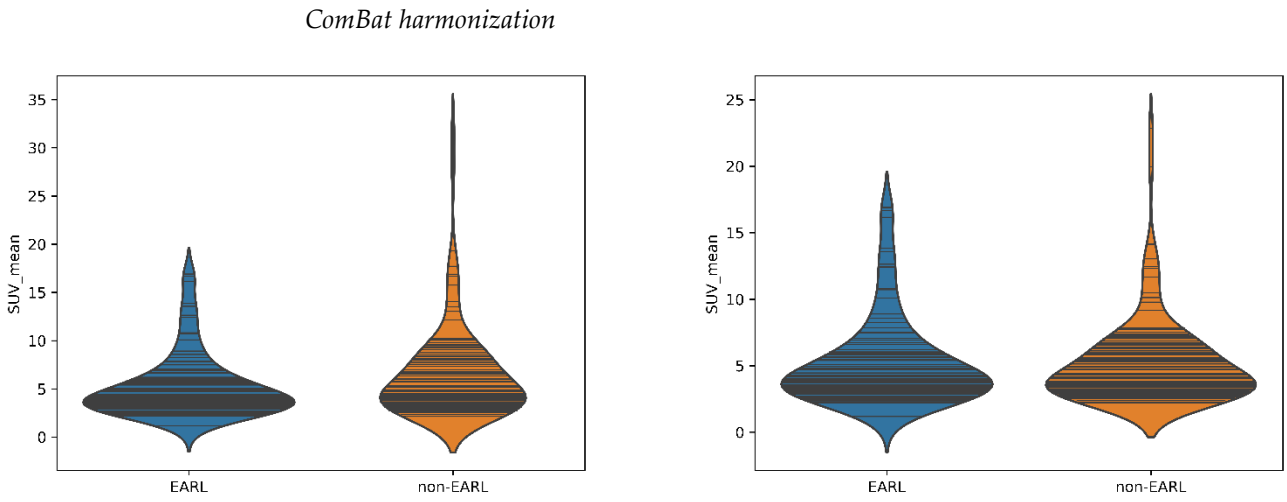


Figure 2. Violin plots of SUV_{mean} values for EARL-compliant and non-EARL-compliant PET images (a) before and (b) after ComBat harmonization.

SUV_{mean} values significantly differed between EARL-compliant and non-EARL-compliant images (figure 2a, p=0.04). Therefore, ComBat harmonization was performed, resulting in no significant differences in SUV_{mean} (figure 2b, p=0.95). Therefore, the remainder of the manuscript uses the harmonized data.

Neither of the models could identify metastases with AUCs of 0.31, 0.35, and 0.30 for the clinical, radiomic, and clinicoradiomic model, respectively (table 2, figure 3). In the sham experiment, no model yielded a validation AUC different from 0.5 (range: 0.49–0.51), thereby validating the results. Oversampling the dataset to reduce the class imbalance between patients with and without metastases did not improve the classification performance with AUCs of 0.26, 0.37, and 0.31, respectively (table S1, figure S1).

Table 2. The AUC and accuracy of the classifier for the training and validation sets of the complete dataset. The values are given for clinical variables, radiomic features and both combined using a logistic regression classifier.

		Training set		**Validation set**	
		AUC	Accuracy	AUC	Accuracy
Clinical model	Lauren Classification (intestinal/mixed or diffuse-type) Clinical T-stage (T3, T4a or T4b) Degree of Differentiation (well-moderate or poor-undifferentiated) Her2Neu-status (negative or positive) Clinical N-stage (N0 or N+)	0.71	79%	0.31	81%
Radiomics model	Cluster Prominence (GLCM) Autocorrelation (GLCM) Variance (First Order) Large Area High Gray Level Emphasis (GLSZM) Surface Volume Ratio (Shape) Coarseness (NGTDM) Inverse Difference Normalized (GLCM)	0.66	79%	0.35	80%

	Small Dependence Low Gray Level Emphasis (GLDM) Size Zone Non Uniformity Normalized (GLSZM) Large Dependence High Gray Level Emphasis (GLDM)				
Clinicoradiomic model	All variables specified above	0.76	78%	0.30	81%

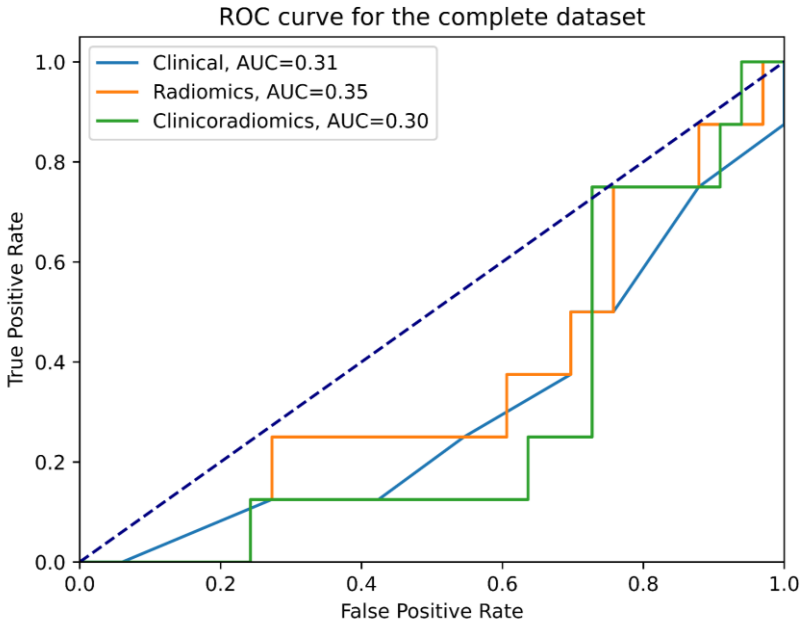


Figure 3. ROC curves of the three models for the validation set. The dashed blue line is the reference line with an AUC of 0.5.

Subgroup analysis based on the Lauren classification did not improve the predictive performance of the investigated models. In intestinal and mixed-type tumors, neither of the models could identify metastases with AUCs of 0.60, 0.35, and 0.33 for the clinical, radiomic, and clinicoradiomic model, respectively (table 3, figure 4). In the sham experiment, no model yielded a validation AUC different from 0.5 (range: 0.5–0.51).

Table 3. The AUC and accuracy of the classifier for the validation and training set of the intestinal- and mixed-type tumors. The values are given for clinical variables, radiomic features and when both are combined using a logistic regression classifier.

		Training set		**Validation set**	
		AUC	Accuracy	AUC	Accuracy
Clinical model	Clinical T-stage (T3, T4a or T4b) Degree of Differentiation (well-moderate or poor-undifferentiated) Her2Neu-status (negative or positive) Clinical N-stage (N0 or N+)	0.68	82%	0.60	80%
Radiomics model	Skewness (First Order) Small Dependence Low Gray Level Emphasis (GLDM) Flatness (Shape) Gray Level Non Uniformity (GLSZM) Informational Measure of Correlation (IMC) 2 (GLCM)	0.84	87%	0.35	84%

	Elongation (Shape)				
Clinicoradiomic model	All variables specified above	0.87	89%	0.33	80%

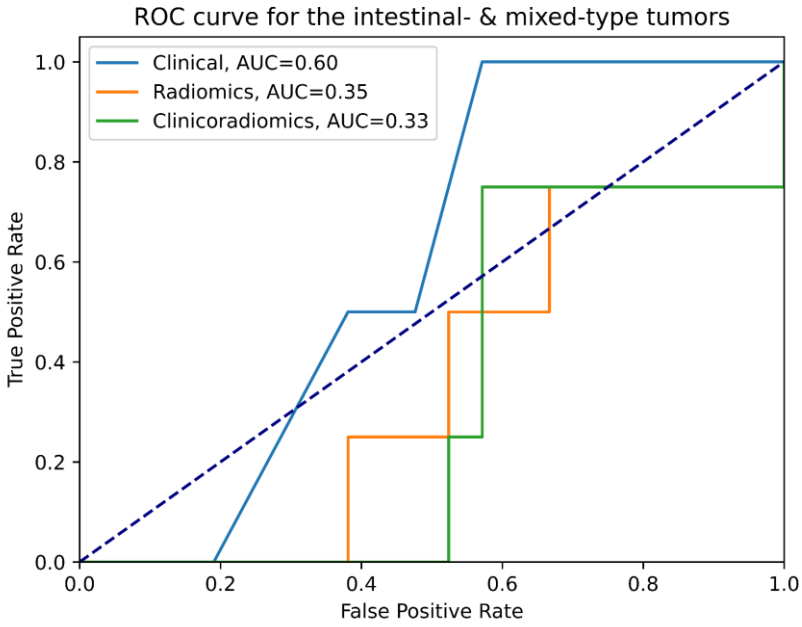


Figure 4. ROC curves of the three models for the validation set of intestinal- and mixed-type tumors. The dashed blue line is the reference line with an AUC of 0.5.

In diffuse tumors, no model could identify metastases, with AUCs of 0.56, 0.51 and 0.49 for the clinical, radiomic, and clinicoradiomic model, respectively (table 4, figure 5). In the sham experiment, no model yielded a validation AUC different from 0.5 (range: 0.49–0.5).

Table 4. The AUC and accuracy of the classifier for the validation and training set of the diffuse-type tumors. The values are given for clinical variables, radiomic features and when both are combined using a logistic regression classifier.

		Training set		**Validation set**	
		AUC	Accuracy	AUC	Accuracy
Clinical model	Clinical T-stage (T3, T4a or T4b) Degree of Differentiation (well-moderate or poor-undifferentiated) Her2Neu-status (negative or positive) Clinical N-stage (N0 or N+)	0.60	77%	0.56	81%
Radiomics model	Uniformity (First Order) Coarseness (NGTDM) Small Area Emphasis (GLSZM) Small Area Low Gray Level Emphasis (GLSZM)	0.89	83%	0.51	88%
Clinicoradiomic model	All variables specified above	0.94	85%	0.49	81%

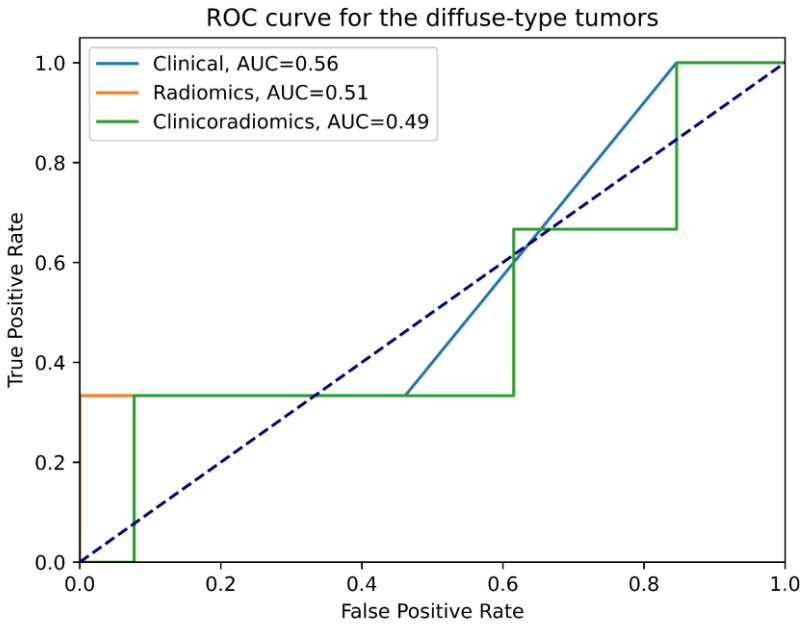


Figure 5. ROC curves of the three models for the validation set of diffuse-type tumors. The dashed blue line is the reference line with an AUC of 0.5.

4. Discussion

In this multicenter study, we built and validated an [^{18}F]FDG-PET radiomics model to preoperatively identify peritoneal and distant metastases in 206 patients with surgically resectable, advanced gastric adenocarcinoma (cT3-4b, N0-3, M0). However, we found that neither the radiomics model nor radiomics combined with clinical variables showed any added value in identifying distant metastases. In addition, subgroup analyses based on the Lauren classification did not identify a tumour subtype that could benefit from [^{18}F]FDG-PET radiomics. In addition to the PLASTIC study, which did not find added value in the qualitative assessment of [^{18}F]FDG-PET/CT in gastric cancer, this study shows that quantitative assessment does not have added value either.

In the complete dataset, none of the investigated models could predict the presence of metastases. Contrarily, Liu et al. [13] found that distant metastases could be correctly predicted by a PET radiomics model (accuracy: 85.2%) in 355 patients with gastric cancer who underwent total or radical gastrectomy, while some metastases in the validation cohort were incorrectly predicted by the visual assessment of the [^{18}F]FDG-PET/CT scan (accuracy: 55%). Furthermore, Xue et al. [14] compared a clinical model, an [^{18}F]FDG-PET-based radiomics model, and a clinicoradiomic model to predict peritoneal metastases in gastric cancer and found in a single-centred study that all models can predict these metastases with validation AUCs of 0.87, 0.69 and 0.90, respectively. However, fewer radiomic features were extracted, fewer clinical parameters were implemented in the model, and only peritoneal metastases were predicted. Another reason for the low performance of the clinical and clinicoradiomic model in our study might be that only 21% of the patients had metastases in comparison to 31% in the study by Xue et al.. Therefore in our study, the model was trained on fewer events, possibly making our model less robust. In addition, the class imbalance in our dataset complicated the classification task since a machine learning model might focus on learning the characteristics of the majority class while neglecting the characteristics of the minority class, i.e., the presence of metastases, which are of more interest. Class imbalance might be solved using oversampling techniques, a form of data augmentation that synthesizes new examples of the minority class [29].

In contrast to the study of Lv et al. [30] in lung adenocarcinoma, oversampling of the minority class in our training set did not result in higher AUCs for all three models (results

shown in table S1 and figure S1). Please note that oversampling will result in a distribution with half of the patient population with metastases, which might improve the classification task. However, the target class is based on only a small sample of patients, which might reduce the model's generalizability [31]. Also, from a clinical perspective, oversampling might be unrealistic. In our study, oversampling had no additional value, indicating that increasing the number of metastases in our dataset would not improve the model performance. Therefore, radiomics showed no added value in our cohort.

Contrary to our expectations, the subgroup analyses based on the Lauren classification showed neither model could predict metastases better in intestinal/mixed type or diffuse-type tumours. Since intestinal type and diffuse type tumours show different metastatic patterns and [^{18}F]FDG-uptake [32], it was hypothesized that [^{18}F]FDG-PET radiomics could identify metastases in at least a specific subgroup of patients. Diffuse-type tumours more frequently present with peritoneal metastases, while intestinal-type tumours more often result in other distant metastases (i.e. liver, lung) [26]. Also, according to a study by Kim et al., intestinal-type tumours showed significantly higher [^{18}F]FDG-uptake compared to diffuse-type tumours [32]. Furthermore, [^{18}F]FDG-PET/CT has shown a higher sensitivity for detecting recurrence in gastric cancer in [^{18}F]FDG-avid primary tumours compared to non-[^{18}F]FDG-avid tumours [33]. Therefore, it was hypothesized that radiomics performs better in [^{18}F]FDG-avid tumours (such as intestinal-type tumours) than in [^{18}F]FDG-non-avid tumours since, in the case of a fixed bin width, the larger range of voxel values within the VOI enables more variation in the values of some texture features. Nevertheless, for the intestinal/mixed type features, the shape features elongation and flatness demonstrated significant differences in values between patients with and without metastases, but this did not result in better model performances. Shape features are, to a lesser extent, affected by [^{18}F]FDG avidity than intensity and texture features.

A study by Wang et al. [11] showed that a Lauren classification-based model on CT-based radiomics can predict lymph node metastases in gastric cancer (validation AUC 0.84). However, to the best of our knowledge, no other study has been conducted on PET-based radiomics for distant metastases yet. Our subgroup analysis based on the Lauren classification showed that PET-based radiomics could not predict distant or peritoneal metastases in intestinal and mixed-type tumours or diffuse-type tumours in our cohort.

The AUCs of the training sets are much higher than those of the validation set, which may indicate that the model is overfitting due to many features compared to a relatively small number of patients [34]. Due to the large number of features, the model describes random error in the data rather than relationships between variables. This seems especially the case in the radiomics and clinicoradiomic model. Further analysis is needed to understand how to tune the algorithms more accurately, decreasing overfitting.

All analyses have been validated by sham experiments. It is advised to perform such sham experiments to ensure the results are present in the data and do not rely on coincidence [28]. No model yielded an AUC different from 0.5 in these sham experiments, validating the algorithms.

There are some limitations to this research. The data were collected from sixteen healthcare institutes in the Netherlands. The advantage of this is a larger patient population which is representative of daily clinical practice. However, it also resulted in [^{18}F]FDG-PET/CT scans made by different scanners and different reconstruction protocols, as is often the case in clinical practice. Although acquisition of the [^{18}F]FDG-PET/CT was performed following the EANM guidelines, reconstructions were not all performed according to the EARL ^{18}F standard 1, increasing variability and reducing repeatability and reproducibility of the extracted radiomic features [35]. To minimize the difference between EARL-compliant images and images reconstructed with site-specific reconstruction protocols, ComBat harmonization towards the EARL-compliant scans has been performed in the current study, so the impact on the repeatability and reproducibility of the extracted features is expected to be minimal.

Furthermore, for three of the eight clinical variables, more than 20% of the values were unknown and imputed based on fourteen other clinical variables (Appendix 3). This might lead to a biased outcome since the imputed values were substituted based on available data, assuming that these are similar to the ones present in the dataset. According to research by Jakobsen et al. [36], variables with 40% or more missing values should be discarded. In our data set, this would mean discarding over 60% of the data, thereby decreasing the data set substantially. Since the missing values were at random, the choice has been made to include all data and impute missing variables.

5. Conclusions

Quantitative [^{18}F]FDG PET assessment using radiomics did not contribute to the preoperative identification of distant and peritoneal metastases in patients with surgically resectable, locally advanced gastric adenocarcinoma (cT3-4b, N0-3, M0) in a large Dutch multicentric cohort.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, **Collaborators PLASTIC Study Group**, **Table S1.** The AUC and accuracy of the classifier for the validation and training set of the oversampled dataset. The values are given for clinical variables, radiomic features and when both are combined using a logistic regression classifier. **Figure S1.** ROC curves of the three models for the validation set of the oversampled dataset. The dashed blue line is the reference line with an AUC of 0.5.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021 May;71(3):209-249. Doi: 10.3322/caac.21660. Epub 2021 Feb 4. PMID: 33538338.
2. Cunningham D, Allum WH, Stenning SP, Thompson JN, Van de Velde CJ, Nicolson M, Scarffe JH, Lofts FJ, Falk SJ, Iveson TJ, Smith DB, Langley RE, Verma M, Weeden S, Chua YJ, MAGIC Trial Participants. Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer. N Engl J Med. 2006 Jul 6;355(1):11-20. Doi: 10.1056/NEJMoa055531. PMID: 16822992.
3. Al-Batran SE, Homann N, Pauligk C, Goetze TO, Meiler J, Kasper S, Kopp HG, Mayer F, Haag GM, Luley K, Lindig U, Schmiegel W, Pohl M, Stoehlmacher J, Folprecht G, Probst S, Prasnikar N, Fischbach W, Mahlberg R, Trojan J, Koenigsmann M, Martens UM, Thuss-Patience P, Egger M, Block A, Heinemann V, Illerhaus G, Moehler M, Schenk M, Kullmann F, Behringer DM, Heike M, Pink D, Teschendorf C, Löhr C, Bernhard H, Schuch G, Rethwisch V, von Weikersthal LF, Hartmann JT, Kneba M, Daum S, Schulmann K, Weniger J, Belle S, Gaiser T, Oduncu FS, Güntner M, Hozaeel W, Reichart A, Jäger E, Kraus T, Mönig S, Bechstein WO, Schuler M, Schmalenberg H, Hofheinz RD; FLOT4-AIO Investigators. Perioperative chemotherapy with fluorouracil plus leucovorin, oxaliplatin, and docetaxel versus fluorouracil or capecitabine plus cisplatin and epirubicin for locally advanced, resectable gastric or gastro-oesophageal junction adenocarcinoma (FLOT4): a randomized, phase 2/3 trial. Lancet. 2019 May 11;393(10184):1948-1957. Doi: 10.1016/S0140-6736(18)32557-1. Epub 2019 Apr 11. PMID: 30982686.
4. Integraal Kankercentrum Nederland. Diagnostiek, slokdarm- en maagkanker in Nederland. Published 2021. Accessed October 10th, 2022. https://iknl.nl/getmedia/196f17c1-3c86-41b8-ad9d-721ab0ba81d8/slokdarm_maag_def.pdf
5. Wang Z, Chen JQ. Imaging in assessing hepatic and peritoneal metastases of gastric cancer: a systematic review. BMC Gastroenterol 2011;11:19-230X-11-19.
6. Brenkman HJF, Gertsen EC, Vegt E, et al. Evaluation of PET and laparoscopy in STagIng advanced gastric cancer: A multicenter prospective study (PLASTIC-study). BMC Cancer. 2018;18(1):450.
7. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology. 2016 Feb;278(2):563-77. doi: 10.1148/radiol.2015151169. Epub 2015 Nov 18. PMID: 26579733; PMCID: PMC4734157.
8. Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P, Cook G. Introduction to Radiomics, Journal of Nuclear Medicine April 2020, 61 (4) 488-495; DOI: https://doi.org/10.2967/jnumed.118.222893
9. Noortman WA, Vriens D, Grootjans W, Tao Q, de Geus-Oei LF, Van Velden FH. Nuclear medicine radiomics in precision medicine: why we can't do without artificial intelligence. Q J Nucl Med Mol Imaging. 2020 Sep;64(3):278-290. Doi: 10.23736/S1824-4785.20.03263-X. Epub 2020 May 12. PMID: 32397702.
10. Dong D, Fang MJ, Tang L, Shan XH, Gao JB, Giganti F, Wang RP, Chen X, Wang XX, Palumbo D, Fu J, Li WC, Li J, Zhong LZ, De Cobelli F, Ji JF, Liu ZY, Tian J. Deep learning radiomic nomogram can predict the number of lymph node metastasis in locally advanced gastric cancer: an international multicenter study. Ann Oncol. 2020 Jul;31(7):912-920. Doi: 10.1016/j.annonc.2020.04.003. Epub 2020 Apr 15. PMID: 32304748.
11. Wang Y, Liu W, Yu Y, Liu JJ, Xue HD, Qi YF, Lei J, Yu JC, Jin ZY. CT radiomics nomogram for the preoperative prediction of lymph node metastasis in gastric cancer. Eur Radiol. 2020 Feb;30(2):976-986. Doi: 10.1007/s00330-019-06398-z. Epub 2019 Aug 29. PMID: 31468157.
12. Dong D, Tang L, Li ZY, Fang MJ, Gao JB, Shan XH, Ying XJ, Sun YS, Fu J, Wang XX, Li LM, Li ZH, Zhang DF, Zhang Y, Li ZM, Shan F, Bu ZD, Tian J, Ji JF. Development and validation of an individualized nomogram to identify occult peritoneal metastasis in patients with advanced gastric cancer. Ann Oncol. 2019 Mar 1;30(3):431-438. Doi: 10.1093/annonc/mdz001. PMID: 30689702;
13. Liu Q, Li J, Xin B, Sun Y, Feng D, Fulham MJ, Wang X, Song S. 18F-FDG PET/CT Radiomics for Preoperative Prediction of Lymph Node Metastases and Nodal Staging in Gastric Cancer. Front Oncol. 2021 Sep 13;11:723345. Doi: 10.3389/fonc.2021.723345. PMID: 34589429; PMCID: PMC8474469.
14. Xue B, Jiang J, Chen L, Wu S, Zheng X, Zheng X and Tang K (2021) Development and Validation of a Radiomics Model Based on 18F-FDG PET of Primary Gastric Cancer for Predicting Peritoneal Metastasis. Front. Oncol. 11:740111. Doi: 10.3389/fonc.2021.740111
15. Gertsen EC, Brenkman HJF, van Hillegersberg R, et al. 18F-Fludeoxyglucose–Positron Emission Tomography/Computed Tomography and Laparoscopy for Staging of Locally Advanced Gastric Cancer: A Multicenter Prospective Dutch Cohort Study (PLASTIC). JAMA Surg. 2021;156(12):e215340. Doi:10.1001/jamasurg.2021.5340
16. Ulrich et al., UIowa QIN PET volume indices measurement tool for 3D Slicer, assessed on 30-11-2022 via https://github.com/QIICR/PET-IndiC
17. Ulrich et al., 3D Slicer extension containing tools for importing DICOM PET series, assessed on 30-11-2022 via https://github.com/QIICR/Slicer-PETDICOMExtension
18. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. J Nucl Med. 2009 May;50 Suppl 1(Suppl 1):122S-50S. doi: 10.2967/jnumed.108.057307. PMID: 19403881; PMCID: PMC2755245.
19. Frings V, van Velden FH, Velasquez LM, Hayes W, van de Ven PM, Hoekstra OS, Boellaard R. Repeatability of metabolically active tumor volume measurements with FDG PET/CT in advanced gastrointestinal malignancies: a multicenter study. Radiology. 2014 Nov;273(2):539-48. Doi: 10.1148/radiol.14132807. Epub 2014 May 26. PMID: 24865311.
20. van Griethuysen JJM, Fedorov A, Parmar C et al (2017) Computational radiomics system to decode the radiographic phenotype. Cancer Res 77:e104–e107

21. Leijenaar, R. T. H. et al. The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis. Sci. Rep. 5, 11075; doi: 10.1038/srep11075 (2015).
22. Orlhac F, Eertink JJ, Cottereau A, Zijlsttra JM, Thieblemont C, Meignan MA, Boellaard R, Buvat I. A guide to ComBat harmonization of imaging biomarkers in multicenter studies, Journal of Nuclear Medicine September 2021, jnumed.121.262464; DOI: https://doi.org/10.2967/jnumed.121.262464
23. Multi-site harmonization in Python with neuroCombat, https://github.com/Jfortin1/neuroCombat/tree/ac82a067412078680973ddf72bd634d51deae735
24. Wu L, Wang C, Tan X, et al. Radiomics approach for preoperative identification of stages I-II and III-IV of esophageal cancer. Chin J Cancer Res. 2018;30(4):396-405. Doi:10.21147/j.issn.1000-9604.2018.04.02
25. Uddin, S., Khan, A., Hossain, M. et al. Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak 19, 281 (2019). https://doi.org/10.1186/s12911-019-1004-8
26. Lauren P. The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification. Acta Pathol Microbiol Scand. 1965;64:31-49. Doi: 10.1111/apm.1965.64.1.31. PMID: 14320675.
27. Guillaume Lemaî}tre and Fernando Nogueira and Christos K. Aridas, Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning, Journal of Machine Learning Research, 18(17):1−5, 2017 via https://imbalanced-learn.org/stable/references/generated/imblearn.over_sampling.RandomOverSampler.html
28. Buvat I., Orlhac F. The Dark Side of Radiomics: On the Paramount Importance of Publishing Negative Results, Journal of Nuclear Medicine Nov 2019, 60 (11) 1543-1544; DOI: 10.2967/jnumed.119.235325
29. Branco P, Torgo L, Ribeiro R., A Survey of Predictive Modelling under Imbalanced Distributions, May 2015, https://doi.org/10.48550/arXiv.1505.01658
30. Lv J, Chen X, Liu X, Du D, Lv W, Lu L and Wu H (2022) Imbalanced Data Correction Based PET/CT Radiomics Model for Predicting Lymph Node Metastasis in Clinical Stage T1 Lung Adenocarcinoma. Front. Oncol. 12:788968. Doi: 10.3389/fonc.2022.788968
31. Burnaev E., Erofeev P. Papanov A. Influence of Resampling on Accuracy of Imbalanced Classification, arXiv:1707.03905v1 [stat.ML] 12 Jul 2017
32. Kim HW, Won KS, Song BI, Kang YN. Correlation of Primary Tumor FDG Uptake with Histopathologic Features of Advanced Gastric Cancer. Nucl Med Mol Imaging. 2015 Jun;49(2):135-42. Doi: 10.1007/s13139-015-0327-3. Epub 2015 Feb 25. PMID: 26085859; PMCID: PMC4463880.
33. Kim SJ, Cho YS, Moon SH, Bae JM, Kim S, Choe YS, Kim BT, Lee KH. Primary Tumor ^{18}F-FDG Avidity Affects the Performance of ^{18}F-FDG PET/CT for Detecting Gastric Cancer Recurrence. J Nucl Med. 2016 Apr;57(4):544-50. Doi: 10.2967/jnumed.115.163295. Epub 2015 Dec 17. PMID: 26678615.
34. Wagner, M., Namdar, K., Biswas, A. et al. Radiomics, machine learning, and artificial intelligence—what the neuroradiologist needs to know. Neuroradiology 63, 1957–1967 (2021). https://doi.org/10.1007/s00234-021-02813-9
35. Pfaehler E, Zhovannik I, Wei L, Boellaard R, Dekker A, Monshouwer R, El Naqa I, Bussink J, Gillies R, Wee L, Traverso A. A systematic review and quality of reporting checklist for repeatability and reproducibility of radiomic features. Phys Imaging Radiat Oncol. 2021 Nov 9;20:69-75. doi: 10.1016/j.phro.2021.10.007. PMID: 34816024; PMCID: PMC8591412.
36. Jakobsen, J.C., Gluud, C., Wetterslev, J. et al. When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. BMC Med Res Methodol 17, 162 (2017). https://doi.org/10.1186/s12874-017-0442-1

Supplementary material

Collaborators PLASTIC Study Group
UMCU: Hylke J.F. Brenkman, Frank J. Wessels
AvL: Johanna W. van Sandick
(1) AUMC locatie UvA (2) Cancer Center Amsterdam: Mark I. van Berge Henegouwen, Suzanne S. Gisbertz
(1) AUMC locatie VU (2) Cancer Center Amsterdam: Donald L. van der Peet, Freek Daams
Catharina Ziekenhuis Eindhoven: Misha D.P. Luyer, Grard A.P. Nieuwenhuijzen
Erasmus MC: Sjoerd M. Lagarde, Bas P.L. Wijnhoven, Jan J.B. Van Lanschot
LUMC: Wobbe O. de Steur, Henk H. Hartgrink
Zuyderland MC: Jan H.M.B. Stoot, Karel W.E. Hulsewe
Rijnstate Arnhem: Ernst Jan Spillenaar Bilgen
Almelo ZGT: Marc J. van Det, Ewout A. Kouwenhoven
Elisabeth Twee-Steden Hospital: Joos Heisterkamp
UMCG: Boudewijn van Etten, Jan W. Haveman
Medical Center Leeuwarden: Jean-Pierre Pierie, Hasan H. Eker
Albert Schweizer Dordrecht: Annemieke Y. Thijssen, Eric J.T. Belt
Gelre Apeldoorn: Peter van Duijvendijk, Eelco Wassenaar
ISALA Zwolle: Kevin P. Wevers
Maasstad Hospital, Rotterdam: Lieke Hol

Table S1. The AUC and accuracy of the classifier for the validation and training set of the oversampled dataset. The values are given for clinical variables, radiomic features and when both are combined using a logistic regression classifier.

		Training set		**Validation set**	
		AUC	Accuracy	AUC	Accuracy
Clinical model	Clinical T-stage Degree of Differentiation Her2Neu-status Clinical N-stage	0.57	55%	0.26	39%
Radiomics model	Skewness (First Order) Small Dependence Low Gray Level Emphasis (GLDM) Flatness (Shape) Gray Level Non Uniformity (GLSZM) Informational Measure of Correlation (IMC) 2 (GLCM) Elongation (Shape)	0.65	60%	0.37	24%
Clinicoradiomic model	All variables specified above	0.67	62%	0.31	24%

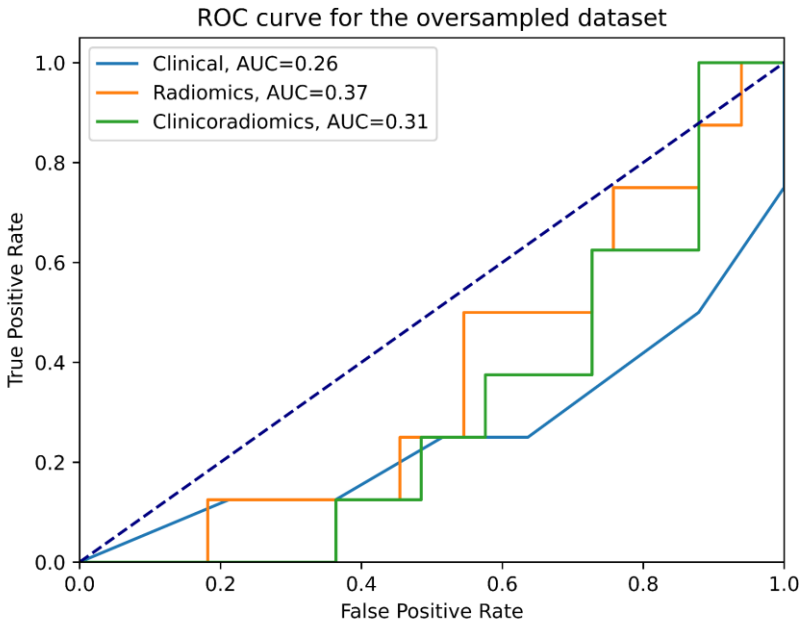


Figure S1. ROC curves of the three models for the validation set of the oversampled dataset. The dashed blue line is the reference line with an AUC of 0.5.

Appendix 1: Overview of extracted radiomic features

Shape

- Elongation
- Flatness
- Least Axis Length
- Major Axis Length
- Maximum 2D Diameter Column
- Maximum 2D Diameter Row
- Maximum 2D Diameter Slice
- Maximum 3D Diameter
- Mesh Volume
- Minor Axis Length
- Sphericity
- Surface Area
- Surface Volume Ratio
- Voxel Volume

First Order

- 10 Percentile
- 90 Percentile
- Energy
- Entropy
- Interquartile Range
- Kurtosis
- Maximum
- Mean Absolute Deviation
- Mean
- Median
- Minimum
- Range
- Robust Mean Absolute Deviation
- Root Mean Squared
- Skewness
- Total Energy
- Uniformity
- Variance

Gray Level Co-Occurrence Matrix (GLCM)

- Autocorrelation
- Joint Average
- Cluster Prominence
- Cluster Shade
- Cluster Tendency
- Contrast
- Correlation
- Difference Average
- Difference Entropy
- Difference Variance
- Joint Energy
- Joint Entropy
- Informational Measure of Correlation (IMC) 1
- Informational Measure of Correlation (IMC) 2
- Inverse Difference Moment
- Inverse Difference Moment Normalized
- Inverse Difference

- Inverse Difference Normalized
- Inverse Variance
- Maximum Probability
- Sum Entropy
- Sum Squares

Grayscale Run-Length Matrices (GLRLM)
- Gray Level Non Uniformity
- Gray Level Non Uniformity Normalized
- Gray Level Variance
- High Gray Level Run Emphasis
- Long Run Emphasis
- Long Run High Gray Level Emphasis
- Long Run Low Gray Level Emphasis
- Low Gray Level Run Emphasis
- Run Entropy
- Run Length Non Uniformity
- Run Length Non Uniformity Normalized
- Run Percentage
- Run Variance
- Short Run Emphasis
- Short Run High Gray Level Emphasis
- Short Run Low Gray Level Emphasis

Gray Level Size Zone Matrix (GLSZM)
- Gray Level Non Uniformity
- Gray Level Non Uniformity Normalized
- Gray Level Variance
- High Gray Level Zone Emphasis
- Large Area Emphasis
- Large Area High Gray Level Emphasis
- Large Area Low Gray Level Emphasis
- Low Gray Level Zone Emphasis
- Size Zone Non Uniformity
- Size Zone Non Uniformity Normalized
- Small Area Emphasis
- Small Area High Gray Level Emphasis
- Small Area Low Gray Level Emphasis
- Zone Entropy
- Zone Percentage
- Zone Variance

Gray Level Dependence Matrix (GLDM)
- Dependence Entropy
- Dependence Non Uniformity
- Dependence Non Uniformity Normalized
- Dependence Variance
- Gray Level Non Uniformity
- Gray Level Variance
- High Gray Level Emphasis
- Large Dependence Emphasis
- Large Dependence High Gray Level Emphasis
- Large Dependence Low Gray Level Emphasis
- Low Gray Level Emphasis
- Small Dependence Emphasis

- Small Dependence High Gray Level Emphasis
- Small Dependence Low Gray Level Emphasis

Neighboring Gray Tone Difference Matrix (NGTDM)

- Busyness
- Coarseness
- Complexity
- Contrast
- Strength

Appendix 2: Overview of ground variables for imputation

- PET positive
- Diagnostic laparoscopy positive
- Primary tumor positive on PET
- Lymph node positive on PET
- Fluid on diagnostic laparoscopy
- Curative or palliative treatment
- Gastric resection
- Curative treatment plan
- Type of resection
- Type of treatment
- Scheme chemotherapy
- Recurrence after six months
- Location of recurrence after six months
- Alive after six months

Appendix 3: Statistical analysis of clinical variables and radiomic features for complete dataset and subgroup analysis

Table 1A. P-values of Mann Whitney U-test for clinical variables and radiomic features included in the complete dataset. $P < 0.05$ is considered significant, given in green.

Clinical variables	**p-value**
Lauren Classification	0.46
Clinical T-stage	0.01
Degree of Differentiation	0.14
Her2Neu-status	0.84
Clinical N-stage	0.03
Radiomic features	**p-value**
Cluster Prominence (GLCM)	0.15
Autocorrelation (GLCM)	0.87
Variance (First Order)	0.11
Large Area High Gray Level Emphasis (GLSZM)	0.07
Surface Volume Ratio (Shape)	0.75
Coarseness (NGTDM)	0.66
Inverse Difference Normalized (GLCM)	0.22
Small Dependence Low Gray Level Emphasis (GLDM)	0.78
Size Zone Non Uniformity Normalized (GLSZM)	0.52
Large Dependence High Gray Level Emphasis (GLDM)	0.82

Table 1B. P-values of Mann Whitney U-test for clinical variables and radiomic features included in the subgroup analysis for intestinal- and mixed-type tumors. $P < 0.05$ is considered significant, given in green.

Clinical variables	**p-value**
Clinical T-stage	0.03
Degree of Differentiation	0.01
Her2Neu-status	0.82
Clinical N-stage	0.02
Radiomic features	**p-value**
Skewness (First Order)	0.04
Small Dependence Low Gray Level Emphasis (GLDM)	0.31
Flatness (Shape)	0.79
Gray Level Non Uniformity (GLSZM)	0.61
Informational Measure of Correlation (IMC) 2 (GLCM)	0.23
Elongation (Shape)	0.01

Table 1C. P-values of Mann Whitney U-test for clinical variables and radiomic features included in the subgroup analysis for diffuse-type tumors. Neither variables or features showed significance.

Clinical variables	**p-value**
Clinical T-stage	0.07
Degree of Differentiation	0.23
Her2Neu-status	0.08
Clinical N-stage	0.35
Radiomic features	**p-value**
Uniformity (First Order)	0.98
Coarseness (NDGTDM)	0.78
Small Area Emphasis (GLSZM)	0.87
Small Area Low Gray Level Emphasis (GLSZM)	0.06

GENERAL DISCUSSION

Patient inclusion, image acquisition, reconstruction & VOI segmentation

The data from this research comes from a multi-centre study. Next to the previously mentioned limitation with EARL and non-EARL scanned scans, other issues arose when using the PET scans. Some scans had to be excluded since they missed DICOM tags or only fused PET/CT scans were available. In most cases, the missing DICOM tags were *PatientWeight* and *SeriesTime*. The *SeriesTime* was implemented manually, but the *PatientWeight* caused some issues. Since the data was only available anonymized, healthcare institutes were unable to find the matching patient numbers and, therefore unable to find the weight. The *PatientWeight* is particularly important since the SUV measurements are scaled to the patient's weight.

To prevent this from happening, a document has been set up with requirements that scans must meet to be considered for inclusion. The document can be found in Appendix A. A small number of patients (n=8) also had to be excluded because all clinical variables were missing. Moreover, the subgroup analysis for the Lauren classification has been done with imputed data where the Lauren classification, degree of differentiation and Her2Neu-status was imputed to increase the number of subjects per subgroup. This might create a bias for the Lauren classification.

Furthermore, the VOI segmentation has been done by only one assessor. Scans with doubtful VOIs have been discussed with a nuclear medicine physician, but scans that the assessor does not doubt are automatically assumed to be true. Since the assessor has little experience with lesion detection, there might be some differences when compared to the segmentation of an experienced assessor. However, since the adaptive threshold is used, it is assumed that these differences will be minor.

Radiomics pipeline

In total, 105 radiomic features are extracted from the VOIs. In the field of radiomics, over 5000 radiomic features are known to this day[41]. Only a fraction of the available radiomic features has been extracted, yielding only a small amount of information. Including other features can lead to more optimal features being selected and may increase the models' performance in detecting metastases. However, it is not the case that more features should be included. The more features included in the model, the higher the chances of overfitting due to the curse of dimensionality. The balance between the number of features and the size of the data set is fragile and an interesting topic for future research.

In supervised feature selection, labelled data sets with a known outcome measure are used for feature selection. With unsupervised feature selection or dimensionality reduction, the outcome is unknown. The algorithm will process unlabeled data and structure it into clusters or relationship patterns. Supervised methods have advantages and disadvantages over unsuper-

vised methods for PET radiomics. Features can be selected more specifically with the desired outcome, but manual feature selection may only capture some of the underlying imaging information. The latter is solved with unsupervised methods, which may result in finding unknown patterns in the data. Unsupervised learning is generally more robust, but the selection of the features and training of the model is without human intervention[42]. A study by Shakir et al. regarding supervised and unsupervised CT radiomics for different types of cancers found that unsupervised methods outperformed the supervised methods in diagnosing the cancer type[43]. To our knowledge, no study has yet validated the differences between supervised and unsupervised PET radiomics.

LASSO is used to select the features implemented in the model from the clustering. LASSO directly sorts the features based on their predictive performance. Goméz et al. investigated the use of different supervised machine learning algorithms in combination with classifiers for metabolic response prediction of metastatic breast cancer lesions in PET/CT imaging[44]. It was found that out of seven feature selection - both supervised and unsupervised - algorithms, LASSO predicted the metabolic response the best. When in combination with classifiers, LASSO in combination with SVM (AUC = 0.93 ± 0.06) followed by LASSO in combination with RF (AUC = 0.92 ± 0.03). LASSO is a commonly used feature selection algorithm for PET radiomics[32] due to its value use in discarding unimportant features.

Originally, random forest was used as a classifying algorithm due to its positive reputation for complex models[45]. The results using this classifier are given in Appendix 2. The main reason for choosing another classifier was the clinical validation parameters of the training set. Even though the results for most models and data sets are better than the used logistic regression algorithm, the AUC and accuracy are remarkably high for all training sets. Classification techniques such as the used random forest draw a uniform random sample from the data. Due to the minority character of the group with metastases, the model tends to be biased for the group without metastases[46]. The no metastases group is often predicted correctly, while the group with metastases is not, resulting in high accuracy, even though the group with metastases is predicted poorly in all models.
Furthermore, the random forest algorithm is overfitting much more than the logistic regression algorithm. Decision trees are known to be more prone to overfitting noise due to the strict boundary character of the decision trees for the training samples[45]. Logistic regression is more flexible since it bases its final prediction on the relationship between all included features without strict decision boundaries. The combination of the minority bias and overfitting leads to higher training clinical validation parameters of random forest compared to logistic regression.

Future perspective

Although PET radiomics showed great results in gastric cancer, future research might lean more towards deep learning and artificial intelligence. Volume segmentation and complex lesions can benefit from (semi-)-automatic segmentation, reducing the time-consuming task of VOI segmentation[47]. Deep learning methods in gastric cancer showed great potential and can unravel patterns that have not been seen before[42]. Since both qualitative and quantitative assessments did not find additional value, more complex algorithms might be. Deep learning models require a large collection of data, hence why deep learning has not been the scope of this study[48]. Furthermore, features often do not have a clear biological or physical model, making the deep learning model an incomprehensible ”black box”.

Lastly, radiomics remains a complex and abstract field of machine learning for most clinicians. It is hard to visualize and does not directly impact the procedure of imaging itself[49]. Therefore,

many steps need to be taken to normalize using radiomics in a clinical setting to make patients benefit from the innovations in the field of radiomics in health care.

REFERENCES

[1] Integraal Kankercentrum Nederland, Incidentie, https://iknl.nl/nkr-cijfers, assessed on 26-10-2022

[2] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021 May;71(3):209-249. doi: 10.3322/caac.21660. Epub 2021 Feb 4. PMID: 33538338.

[3] Morgan E, Arnold M, Camargo MC, Gini A, Kunzmann AT, Matsuda T, Meheus F, Verhoeven RHA, Vignat J, Laversanne M, Ferlay J, Soerjomataram I. The current and future incidence and mortality of gastric cancer in 185 countries, 2020-40: A population-based modelling study. EClinicalMedicine. 2022 Apr 21;47:101404. doi: 10.1016/j.eclinm.2022.101404. PMID: 35497064; PMCID: PMC9046108.

[4] Yumo Xie, Lishuo Shi, Xiaosheng He, Yanxin Luo, Gastrointestinal cancers in China, the USA, and Europe, Gastroenterology Report, Volume 9, Issue 2, April 2021, Pages 91–104, https://doi.org/10.1093/gastro/goab010

[5] Cho E, Kang MH, Choi KS, Suh M, Jun JK, Park EC. Cost-effectiveness outcomes of the national gastric cancer screening program in South Korea. Asian Pac J Cancer Prev. 2013;14(4):2533-40. doi: 10.7314/apjcp.2013.14.5.2533. PMID: 23725170.

[6] Hamashima C; Systematic Review Group and Guideline Development Group for Gastric Cancer Screening Guidelines. Update version of the Japanese Guidelines for Gastric Cancer Screening. Jpn J Clin Oncol. 2018 Jul 1;48(7):673-683. doi: 10.1093/jjco/hyy077. PMID: 29889263.

[7] Smyth EC, Nilsson M, Grabsch HI, van Grieken NC, Lordick F. Gastric cancer. Lancet. 2020 Aug 29;396(10251):635-648. doi: 10.1016/S0140-6736(20)31288-5. PMID: 32861308.

[8] Smyth E, Schöder H, Strong VE, et al. A prospective evaluation of the utility of 2-deoxy-2-[(18) F]fluoro-D-glucose positron emission tomography and computed tomography in staging locally advanced gastric cancer. Cancer 2012; 118: 5481–88.

[9] Terese Winslow. https://www.cancer.gov/types/stomach/patient/stomach-treatment-pdq, assessed on 24-5-2022.

[10] Brierly J, Gospodarowicz MK, Wittekind C, TNM Classification of Malignant Tumours, 8th edition, 2016 Nov, ISBN: 978-1-119-26356-2.

[11] Lauren P. The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification. Acta Pathol Microbiol Scand. 1965;64:31-49. doi: 10.1111/apm.1965.64.1.31. PMID: 14320675.

[12] Ma J, Shen H, Kapesa L, Zeng S. Lauren classification and individualized chemotherapy in gastric cancer. Oncol Lett. 2016;11(5):2959-2964. doi:10.3892/ol.2016.4337

[13] Desai JP, Moustarah F. Peritoneal Metastasis. StatPearls, StatPearls Publishing; 2022 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK541114/, assessed on 03-06-2022

[14] Gill RS, Al-Adra DP, Nagendran J, Campbell S, Shi X, Haase E, Schiller D. Treatment of gastric cancer with peritoneal carcinomatosis by cytoreductive surgery and HIPEC: a systematic review of survival, mortality, and morbidity. J Surg Oncol. 2011 Nov 01;104(6):692-8.

[15] Zhang Y, Lin Y, Duan J, Xu K, Mao M, Wang X. A Population-Based Analysis of Distant Metastasis in Stage IV Gastric Cancer. Med Sci Monit. 2020;26:e923867. Published 2020 May 15. doi:10.12659/MSM.923867

[16] Brennan MF, Karpeh MS: Surgery for gastric cancer: the American view. Semin Oncol 23 (3): 352-9, 1996.

[17] Integraal Kankercentrum Nederland. Diagnostiek, behandeling en follow-up van het maagcarcinoom 2016. Published 2017. Accessed September 20, 2021. https://richtlijnendatabase.nl/ gerelateerde_documenten/f/16316/IKNL%20richtlijn%20Maagcarcinoom.pdf

[18] Prince, J. L., Links, J. M. (2006). Medical Imaging Signals and Systems. Upper Saddle River, N.J: Pearson Prentice Hall.

[19] Jiang W, Chalich Y, Deen MJ. Sensors for Positron Emission Tomography Applications. Sensors. 2019; 19(22):5019. https://doi.org/10.3390/s19225019

[20] Cherry S., R. Sorenson J. A. Phelps M. E. (2012). Physics in nuclear medicine (4th ed.). Elsevier/Saunders. Retrieved November 2 2022 from http://www.clinicalkey.com/dura/browse/bookChapter/3-s2.0-C20090516352.

[21] 18F-FDG PET/CT in Oncology, Summary of EANM guidelines, https://richtlijnendatabase.nl/gerelateerde_documenten/f/17259/18F-FDG%20PETCT%20in%20Oncology.pdf, assessed on 24-5-2022

[22] Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P, Cook G. Introduction to Radiomics, Journal of Nuclear Medicine April 2020, 61 (4) 488-495; DOI: https://doi.org/10.2967/jnumed.118.222893

[23] Langs, G., Röhrich, S., Hofmanninger, J. et al. Machine learning: from radiomics to discovery and routine. Radiologe 58, 1–6 (2018). https://doi.org/10.1007/s00117-018-0407-3

[24] Noortman WA, Vriens D, Grootjans W, Tao Q, de Geus-Oei LF, Van Velden FH. Nuclear medicine radiomics in precision medicine: why we can't do without artificial intelligence. Q J Nucl Med Mol Imaging. 2020 Sep;64(3):278-290. doi: 10.23736/S1824-4785.20.03263-X. Epub 2020 May 12. PMID: 32397702.

[25] Lucignani, G.a; Paganelli, G.b; Bombardieri, E.c The use of standardized uptake values for assessing FDG uptake with PET in oncology: a clinical perspective, Nuclear Medicine Communications: July 2004 - Volume 25 - Issue 7 - p 651-656 doi: 10.1097/01.mnm.0000134329.30912.49

[26] Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. J Nucl Med. 2009 May;50 Suppl 1(Suppl 1):122S-50S. doi: 10.2967/jnumed.108.057307. PMID: 19403881; PMCID: PMC2755245.

[27] van Griethuysen, J. J. M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R. G. H., Fillon-Robin, J. C., Pieper, S., Aerts, H. J. W. L. (2017). Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Research, 77(21), e104–e107. https://doi.org/10.1158/0008-5472.CAN-17-0339

[28] Bellman, Richard Ernest; Rand Corporation (1957). Dynamic programming. Princeton University Press. p. ix. ISBN 978-0-691-07951-6.,

[29] Duin, R.P.; Pekalska, E. Dissimilarity Representation for Pattern Recognition, The: Foundations and Applications; World Scientific: Singapore, 2005; Volume 64.

[30] Altman, N., Krzywinski, M. The curse(s) of dimensionality. Nat Methods 15, 399–400 (2018). https://doi.org/10.1038/s41592-018-0019-x

[31] Wagner, M., Namdar, K., Biswas, A. et al. Radiomics, machine learning, and artificial intelligence—what the neuroradiologist needs to know. Neuroradiology 63, 1957–1967 (2021). https://doi.org/10.1007/s00234-021-02813-9

[32] Wu L, Wang C, Tan X, et al. Radiomics approach for preoperative identification of stages I-II and III-IV of esophageal cancer. Chin J Cancer Res. 2018;30(4):396-405. doi:10.21147/j.issn.1000-9604.2018.04.02

[33] Rizzo, S., Botta, F., Raimondi, S. et al. Radiomics: the facts and the challenges of image analysis. Eur Radiol Exp 2, 36 (2018). https://doi.org/10.1186/s41747-018-0068-z

[34] Uddin, S., Khan, A., Hossain, M. et al. Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak 19, 281 (2019). https://doi.org/10.1186/s12911-019-1004-8

[35] Integraal Kankercentrum Nederland. Diagnostiek, slokdarm- en maagkanker in Nederland. Published 2021. Accessed October 10th, 2022. https://iknl.nl/getmedia/196f17c1-3c86-41b8-ad9d-721ab0ba81d8/slokdarm_maag_def.pdf

[36] Cunningham D, Allum WH, Stenning SP, Thompson JN, Van de Velde CJ, Nicolson M, Scarffe JH, Lofts FJ, Falk SJ, Iveson TJ, Smith DB, Langley RE, Verma M, Weeden S, Chua YJ, MAGIC Trial Participants. Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer. N Engl J Med. 2006 Jul 6;355(1):11-20. doi: 10.1056/NEJMoa055531. PMID: 16822992.

[37] Al-Batran SE, Homann N, Pauligk C, Goetze TO, Meiler J, Kasper S, Kopp HG, Mayer F, Haag GM, Luley K, Lindig U, Schmiegel W, Pohl M, Stoehlmacher J, Folprecht G, Probst S, Prasnikar N, Fischbach W, Mahlberg R, Trojan J, Koenigsmann M, Martens UM, Thuss-Patience P, Egger M, Block A, Heinemann V, Illerhaus G, Moehler M, Schenk M, Kullmann F, Behringer DM, Heike M, Pink D, Teschendorf C, Löhr C, Bernhard H, Schuch G, Rethwisch V, von Weikersthal LF, Hartmann JT, Kneba M, Daum S, Schulmann K, Weniger J, Belle S, Gaiser T, Oduncu FS, Güntner M, Hozaeel W, Reichart A, Jäger E, Kraus T, Mönig S, Bechstein WO, Schuler M, Schmalenberg H, Hofheinz RD; FLOT4-AIO Investigators. Perioperative chemotherapy with fluorouracil plus leucovorin, oxaliplatin, and docetaxel versus fluorouracil or capecitabine plus cisplatin and epirubicin for locally advanced, resectable gastric or gastro-oesophageal junction adenocarcinoma (FLOT4): a randomised, phase 2/3 trial. Lancet. 2019 May 11;393(10184):1948-1957. doi: 10.1016/S0140-6736(18)32557-1. Epub 2019 Apr 11. PMID: 30982686.

[38] Evaluation of PET and Laparoscopy in STagIng advanced gastric Cancer: a multicenter prospective study, PLASTIC Study, Dutch Upper GI Cancer Group, METC 16-633/C, version 1.4, 20 June 2017, assessed via plasticstudie.nl, assessed on 30-5-2022

[39] Wang Z, Chen JQ. Imaging in assessing hepatic and peritoneal metastases of gastric cancer: a systematic review. BMC Gastroenterol 2011;11:19-230X-11-19. .

[40] Gertsen EC, Brenkman HJF, van Hillegersberg R, et al. 18F-Fludeoxyglucose–Positron Emission Tomography/Computed Tomography and Laparoscopy for Staging of Locally Advanced Gastric Cancer: A Multicenter Prospective Dutch Cohort Study (PLASTIC). JAMA Surg. 2021;156(12):e215340. doi:10.1001/jamasurg.2021.5340

[41] Limkin EJ, Sun R, Dercle L, Zacharaki EI, Robert C, Reuze S, et al. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. Ann Oncol. 2017;28(6):1191- 206.

[42] Avanzo M, Wei L, Stancanello J, Vallières M, Rao A, Morin O, Mattonen SA, El Naqa I. Machine and deep learning methods for radiomics. Med Phys. 2020 Jun;47(5):e185-e202. doi: 10.1002/mp.13678. PMID: 32418336; PMCID: PMC8965689.

[43] Shakir, H., Deng, Y., Rasheed, H. et al. Radiomics based likelihood functions for cancer diagnosis. Sci Rep 9, 9501 (2019). https://doi.org/10.1038/s41598-019-45053-x

[44] Gómez, O.V., Herraiz, J.L., Udías, J.M., Haug, A., Papp, L., Cioni, D., Neri, E. Analysis of Cross-Combinations of Feature Selection and Machine-Learning Classification Methods Based on [18F]F-FDG PET/CT Radiomic Features for Metabolic Response Prediction of Metastatic Breast Cancer Lesions. Cancers 2022, 14, 2922. https://doi.org/10.3390/ cancers14122922)

[45] Kirasich, Kaitlin; Smith, Trace; and Sadler, Bivin (2018) "Random Forest vs Logistic Regression: Binary Classification for Heterogeneous Datasets," SMU Data Science Review: Vol. 1: No. 3, Article 9. Available at: https://scholar.smu.edu/datasciencereview/vol1/iss3/9

[46] Burnaev E., Erofeev P. Papanov A. Influence of Resampling on Accuracy of Imbalanced Classification, arXiv:1707.03905v1 [stat.ML] 12 Jul 2017

[47] Hatt M., Cheze Le Rest C, Antonorsi N., Tixier F., Tankyevych O., Jaouen V., Lucia F., Bourbonne V., Schick U., Badic B., Visvikis D., Radiomics in PET/CT: Current Status and Future AI-Based Evolutions, Seminars in Nuclear Medicine, Volume 51, Issue 2, 2021, Pages 126-133, ISSN 0001-2998, https://doi.org/10.1053/j.semnuclmed.2020.09.002.

[48] Orlhac F, Nioche C, Klyuzhin I, Rahmim A, Buvat I. Radiomics in PET Imaging:: A Practical Guide for Newcomers. PET Clin. 2021 Oct;16(4):597-612. doi: 10.1016/j.cpet.2021.06.007. PMID: 34537132.

[49] Hustinx, R. Physician centred imaging interpretation is dying out — why should I be a nuclear medicine physician?. Eur J Nucl Med Mol Imaging 46, 2708–2714 (2019). https://doi.org/10.1007/s00259-019-04371-y

A REQUIREMENTS SCANS

Scans to be considered for the radiomics research need to have certain requirements to be feasible for tumor segmentation. The requirements and its importance are based on tumor segmentation in 3D Slicer version 4.11.

What scans are needed?

An *individual* PET scan and *individual* low dose CT scan is necessary. The PET scan is uploaded into 3D Slicer and used for the segmentation itself, the CT scan is used for anatomical congruence between the PET and CT scan. A *fused* scan is not usable, since 3D Slicer cannot demount the PET and CT scan from each other.

What are the requirements of the PET scan?

The most important requirements are the DICOM tags of the PET scan. DICOM tags are metadata elements that are associated with each DICOM image object. These tags are helpful in the organization of the patient images. A user can search for these images based on the data within the DICOM tags.

The required DICOM tags for tumor segmentation in PET scan are the following:

- Units in BQML
- Scale Factor (if Units in CNTS)
- Series Date
- Series Time
- Acquisition Date
- Acquisition Time
- Radiopharmaceutical Start DateTime
- Radiopharmaceutical Information Sequence
- Radionuclide Total Dose
- Patient Weight

What data needs to be saved?

After the VOI is segmented, but before the adaptive threshold is applied, the following data from 3D Slicer needs to be noted:

- SUV_{mean}
- SUV_{median}
- SUV_{min}
- SUV_{max}
- SUV_{peak}
- Volume (mL)

The original PET image is saved as *IMAG.nrrd*, and the VOI segmentation is saved as *PET_VOI_label.nrrd*. Standarizing these variables makes creating the batch input for the feature extraction easier.

B RANDOM FOREST CLASSIFIER

	Training set			Validation set		
	Clinical	**Radiomic**	**Clinicoradiomic**	**Clinical**	**Radiomic**	**Clinicoradiomic**
AUC	0.81	1.00	1.00	0.43	0.56	0.53
Accuracy	84%	98%	99%	65%	76%	76%

Table B.1: The AUC and accuracy of the classifier for the training and validation set of the complete dataset. The values are given for clinical variables, radiomic features and when both are combined using a random forest classifier.

	Training set			Validation set		
	Clinical	**Radiomic**	**Clinicoradiomic**	**Clinical**	**Radiomic**	**Clinicoradiomic**
AUC	0.66	0.94	0.95	0.37	0.70	0.69
Accuracy	60%	88%	88%	43%	67%	66%

Table B.2: The AUC and accuracy of the classifier for the training and validation set of the oversampled dataset. The values are given for clinical variables, radiomic features and when both are combined using a random forest classifier.

	Training set			Validation set		
	Clinical	**Radiomic**	**Clinicoradiomic**	**Clinical**	**Radiomic**	**Clinicoradiomic**
AUC	0.79	1.00	1.00	0.62	0.47	0.53
Accuracy	84%	99%	99%	79%	75%	77%

Table B.3: The AUC and accuracy of the classifier for the training and validation set of the intestinal- and mixed-type tumors. The values are given for clinical variables, radiomic features and when both are combined using a random forest classifier.

	Training set			Validation set		
	Clinical	**Radiomic**	**Clinicoradiomic**	**Clinical**	**Radiomic**	**Clinicoradiomic**
AUC	0.71	1.00	1.00	0.55	0.38	0.40
Accuracy	80%	98%	98%	66%	72%	73%

Table B.4: The AUC and accuracy of the classifier for the training and validation set of the diffuse-type tumors. The values are given for clinical variables, radiomic features and when both are combined using a random forest classifier.

DANKWOORD

Beste lezer,

Je bent nu aangekomen bij het einde van mijn master thesis, wat ook het einde van mijn studententijd in Enschede betekend. Toen ik zeven jaar geleden begon met mijn bachelor had ik nooit kunnen voorspellen hoe mijn leven er uit zou zien gedurende die tijd.
Ten eerste wil ik graag mijn afstudeercommissie bedanken, met in het bijzonder Wyanne Noortman en Dr. de Geus-Oei. Allereerst voor de mogelijkheid om mijn afstudeeropdracht in het LUMC te mogen doen en het vertrouwen dat jullie me gedurende het proces hebben gegeven. Ik heb me nooit 'een student' gevoeld, maar altijd een gelijke. Daarnaast ook dank naar Lianne Triemstra en Cas de Jongh, die altijd een kritische klinische blik op het onderzoek hebben geworpen. Als ingenieur in de dop is makkelijk het klinische aspect uit het oog te verliezen.
Daarnaast wil ik ook graag alle mensen bedanken die ik in Enschede heb leren kennen. Mijn besturen, dispuutsgenoten en alle andere mensen die mijn studententijd hebben gemaakt zoals hij is geweest, het was fantastisch.
Een van mijn grootste danken gaat uit naar de jongens en meiden van afdeling Q in het UMC. Sinds ik in de flexruimte mijn computer heb geclaimd voelde het al een tweede thuis, zeker de afgelopen paar maanden. Ik zal onze wekelijkse taart, karaoke-momenten en dagelijkse bakjes koffie bij de automaat gaan missen.
Maar mijn allergrootste dank gaat uit naar mijn gezin. Mijn vader Erik, moeder Claudia, broertjes Tobias en Camiel en vriend Jotte, bedankt dat jullie mijn rots waren. Ik ben blij dat ik altijd bij jullie terecht kan voor alle overwegingen, vraagstukken, mental breakdowns en verhalen, of dit nou wel of niet studiegerelateerd was. Ik hou van jullie.

Lieke Pullen, december 2022