medRxiv preprint doi: https://doi.org/10.1101/2021.08.19.21262238; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The WORC* database: MRI and CT scans, segmentations, and clinical labels for 930 patients from six radiomics studies

Martijn P.A. Starmans[a,*], Milea J.M. Timbergen[b,c], Melissa Vos[b,c], Guillaume A. Padmos[a], Dirk J. Grünhagen[b], Cornelis Verhoef[b], Stefan Sleijfer[c], Geert J.L.H. van Leenders[d], Florian E. Buisman[e], Francois E.J.A. Willemssen[a], Bas Groot Koerkamp[e], Lindsay Angus[a,c], Astrid A.M. van der Veldt[a,c], Ana Rajicic[c], Arlette E. Odink[a], Michel Renckens[a], Michail Doukas[d], Rob A. de Man[f], Jan N.M. IJzermans[e], Razvan L. Miclea[g], Peter B. Vermeulen[h], Maarten G. Thomeer[a], Jacob J. Visser[a], Wiro J. Niessen[a,i], Stefan Klein[a]

[a]*Department of Radiology and Nuclear Medicine, Erasmus MC, Rotterdam, the Netherlands*
[b]*Department of Surgical Oncology, Erasmus MC Cancer Institute, Rotterdam, the Netherlands*
[c]*Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, the Netherlands*
[d]*Department of Pathology, Erasmus MC, Rotterdam, the Netherlands*
[e]*Department of Surgery, Erasmus MC Cancer Institute, Rotterdam, the Netherlands*
[f]*Department of Hepatology, Erasmus MC, Rotterdam, the Netherlands*
[g]*Department of Radiology and Nuclear Medicine, Maastricht UMC+, Maastricht, the Netherlands*
[h]*Translational Cancer Research Unit, Department of Oncological Research, Oncology Center, GZA Hospitals Campus Sint-Augustinus and University of Antwerp, Antwerp, Belgium*
[i]*Faculty of Applied Sciences, Delft University of Technology, Delft, the Netherlands*

Abstract

The WORC database consists in total of 930 patients composed of six datasets gathered at the Erasmus MC, consisting of patients with: 1) well-differentiated liposarcoma or lipoma (115 patients); 2) desmoid-type fibromatosis or extremity soft-tissue sarcomas (203 patients); 3) primary solid liver tumors, either malignant (hepatocellular carcinoma or intrahepatic cholangiocarcinoma) or benign (hepatocellular adenoma or focal nodular hyperplasia) (186 patients); 4) gastrointestinal stromal tumors (GISTs) and intra-abdominal gastrointestinal tumors radiologically resembling GISTs (246 patients); 5) colorectal liver metastases (77 patients); and 6) lung metastases of metastatic melanoma (103 patients). For each patient, either a magnetic resonance imaging (MRI) or computed tomography (CT) scan, collected from routine clinical care, one or mul-

*Workflow for Optimal Radiomics Classification

*Corresponding author: Tel.: +31-10-7034044
Email address: m.starmans@erasmusmc.nl (Martijn P.A. Starmans)

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.19.21262238; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

tiple (semi-)automatic lesion segmentations, and ground truth labels from a gold standard (e.g., pathologically proven) are available. All datasets are multicenter imaging datasets, as patients referred to our institute often received imaging at their referring hospital. The dataset can be used to validate or develop radiomics methods, i.e., using machine or deep learning to relate the visual appearance to the ground truth labels, and automatic segmentation methods. See also the research article related to this dataset: Starmans et al., *Reproducible radiomics through automated machine learning validated on twelve clinical applications*, Submitted.

Keywords: Radiomics, Segmentation, Magnetic resonance imaging, Computed tomography, Oncology, Sarcoma, Liver, Metastases

---

Specifications Table

Subject	Medical Imaging
Specific subject area	Routine MRI and CT scans, lesion segmentations, clinical labels of six radiomics studies
Type of data	Medical Imaging data (NIfTI files): MRI data: T1-weighted T2-weighted CT data Medical Imaging metadata (JSON files) Segmentations (NIfTI files) Patient data (Excel files): Age Sex Pathological ground truth (Excel files, subject level variables)
How data were acquired	MRI and CT scans were acquired on 177 different scanners. Age and sex were obtained from patient records. Ground truth data were obtained from a gold standard, mainly by pathological analysis of tumor tissue obtained from either biopsy or resection. An exception was made for "typical" focal nodular hyperplasia (FNH) [1], which was confirmed radiologically. Whole-tumor segmentations were semi-automatically annotated by various observers.
Data format	Raw
Parameters for data collection	MRI and CT scans were acquired with a variety of image acquisition protocols.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.19.21262238; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Description of data collection	Pre-treatment imaging data and ground truth data were retrospectively included at the Erasmus MC from patients with: Well-differentiated liposarcoma or lipoma between 2009 - 2018 Desmoid-type fibromatosis and extremity soft-tissue-sarcoma between 1990 - 2018 Primary solid liver tumors between 2002 - 2018 Gastrointestinal stromal tumors or similar intra-abdominal tumors between 2004 - 2017 Colorectal liver metastases between 2003 - 2015 Lung metastases of melanoma between 2012 - 2018
Data source location	Erasmus MC (University Medical Center) Rotterdam The Netherlands
Data accessibility	Repository name: Health-RI XNAT Data identification number: WORC Direct URL to data: https://xnat.bmia.nl/data/projects/worc Data usage agreement: https://xnat.bmia.nl/data/projects/worc/resources/License/files/WORC_data_license.docx Data downloader: https://doi.org/10.5281/zenodo.5119040
Related research article	Starmans et al., *Reproducible radiomics through automated machine learning validated on twelve clinical applications*, Submitted

1. Value of the Data

- This dataset provides imaging data, outlined lesions, age, sex, and ground truth labels (e.g., diagnosis, genetic mutations, biological characteristics), mostly obtained from pathology, for a large number of patients from six different cancer studies. Publicly sharing imaging data with ground truth labels and segmentations benefits reproducibility, enables external validation, and hence accelerates transition to clinical practice [2, 3, 4]. This dataset has been collected in routine clinical care at multiple centers, thus representing the real-life variability and heterogeneity of the data. For these reasons, this dataset is a valuable resource.
- This dataset will be beneficial for researchers working on computer aided diagnosis for cancer based on imaging, specifically in the areas of liposarcoma, desmoid type-fibromatosis, gastrointestinal stromal lesions, sarcoma, primary liver cancer, (colorectal) liver metastases, and (melanoma) lung metastases.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.19.21262238; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

- This data can be used to validate or develop radiomics methods (i.e., using conventional machine learning or deep learning to relate the visual appearance to the ground truth labels) and automated segmentation methods. For example, the data can be used as a large, heterogenous independent test set, or to increase the size and heterogeneity of train sets for developing new methods.

2. Data Description

The WORC dataset contains 930 patients and is composed of six radiomics studies, coined the Lipo (subsection 2.1), Desmoid (subsection 2.2), Liver (subsection 2.3), GIST (subsection 2.4), CRLM (subsection 2.5), and Melanoma (subsection 2.6) dataset. All datasets were collected at the Erasmus MC, Rotterdam, the Netherlands, but are multicenter imaging datasets, as patients referred to our institute often received imaging at their referring hospital. Example images of each dataset are shown in Figure 1.

For each study, five different sources of data are provided:

1. Routine clinical MRI (Lipo, Desmoid, Liver) or CT (GIST, CRLM, Melanoma) scans
2. Details on the acquisition protocols (subsection 2.7)
3. Lesion segmentations
4. Age and sex
5. Pathological ground truth labels

The data is available on an `XNAT` server; an online platform to store (medical) imaging data in a standardized way, allowing access through both a Graphical User Interface (GUI) and an Application Programming Interface (API) [5]. The datasets for this study are publicly hosted on the Health-RI XNAT [1]. Code to download the data locally, and code to reproduce the experiments from Starmans et al. [6] on these datasets, have been released open-source [7].

For each study, details on the ground truth labels and the data collection are given in the respective subsections. The acquisition protocol details for all studies are described

[1] https://xnat.bmia.nl/data/projects/worc

medRxiv preprint doi: https://doi.org/10.1101/2021.08.19.21262238; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

in subsection 2.7. The scans have been converted from DICOM to NIfTI using the `dcm2niix` toolbox version v1.0.20180518 [8]. For each patient, a single scan is included and provided as NIfTI files named *"image.nii.gz"*. The associated details on the scan acquisition protocol are given in a JSON file named *"metadata.json"*. The corresponding segmentation is given in the NIfTI file *"segmentation.nii.gz"*, where a label of 1 indicates a lesion and a label of 0 indicates background. For the CRLM dataset, multiple segmentations of various lesions made by multiple observers are given, see subsection 2.5. The ground truth pathological labels for all studies are combined in the Excel sheet *"Clinical_data.xlsx"* and as labels on subject level in the `XNAT` project to allow for easier automatic processing.

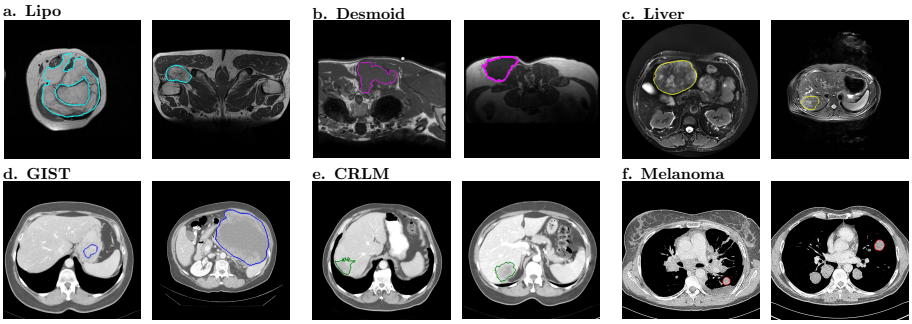

Figure 1: Examples of the 2D slices from the 3D imaging data from the six datasets included in the WORC dataset. For each dataset, for one patient of each of the two classes, the 2D slice in the primary scan direction (e.g., axial) with the largest area of the segmentation is depicted; the boundary of the segmentation is projected in color on the image. The datasets included were from different clinical applications: a. lipomatous tumors [9]; b. desmoid-type fibromatosis [10]; c. primary solid liver tumors [11]; d. gastrointestinal stromal tumors [12]; e. colorectal liver metastases [13]; and f. melanoma [14].

2.1. The Lipo dataset

This dataset consists of 115 patients with either a well-differentiated liposarcoma (WDLPS) (N = 58) or lipoma (N = 58), as described in Vos et al. [9]. One patient has both a WDLPS and a lipoma, thus the dataset in total contains 116 lesions. For each patient, a T1-weighted MRI scan is provided. The ground truth label, i.e., whether a lesion was a WDLPS or lipoma, is represented by the *MDM2* amplification. The

medRxiv preprint doi: https://doi.org/10.1101/2021.08.19.21262238; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

MDM2 amplification status for each patient is provided, where patients have label 1 if the lesion was a WDLPS, and label 0 if the lesion was a lipoma.

For the patient with both a WDLPS and a lipoma, a segmentation is provided for each lesion: *"segmentation_WDLPS.nii.gz"* and *"segmentation_lipoma.nii.gz"*

2.2. The Desmoid dataset

This dataset consists of 203 patients with either desmoid-type fibromatosis (DTF) (N = 72) or extremity soft-tissue sarcomas (STS), i.e, the non-DTF group (N = 131), as described in Timbergen et al. [10]. The non-DTF group consists of 64 myxofibrosarcomas, 31 leiomyosarcomas, and 36 myxoid liposarcomas. For each patient, a T1-weighted MRI scan is provided. The ground truth label, i.e., whether a lesion was a DTF or one of the non-DTF phenotypes, was confirmed by histology. The differential diagnosis for each patient is provided, where patients have label 1 if the lesion was a DTF, and label 0 if the lesion was a non-DTF. The subtype of the non-DTF lesions is also provided.

2.3. The Liver dataset

This dataset consists of 186 patients with either a malignant (N = 94) or benign (N = 93) primary solid liver tumor, as described in Starmans et al. [11]. For each patient, a T2-weighted MRI scan is provided. The malignant group includes 81 hepatocellular carcinoma (HCC) and 13 intrahepatic cholangiocarcinoma (iCCA); the benign group includes 48 hepatocellular adenoma (HCA) and 44 FNH. The ground truth label, i.e., the phenotype of a lesion, was based on pathology. An exception are "typical" FNH [1], for which the ground truth was established radiologically. The differential diagnosis for each patient is provided, where patients have label 1 if the lesion was malignant, and label 0 if the lesion was benign. The phenotype of the lesions is also provided.

2.4. The GIST dataset

This dataset consists of 246 patients with either gastrointestinal stromal lesions (GISTs) (N = 125) or intra-abdominal tumors radiologically resembling GIST (non-GIST) (N = 122), as described in Starmans et al. [12]. One patient has two GISTs, thus

medRxiv preprint doi: https://doi.org/10.1101/2021.08.19.21262238; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

the dataset in total contains 247 lesions. The non-GIST group consists of 22 schwannoma, 25 leiomyosarcoma, 25 leiomyoma, 25 esophageal or gastri junctional adenocarcinoma, and 25 lymphoma. For each patient, a contrast-enhanced venous phase CT scan is provided. The ground truth label, i.e., whether a lesion was a GIST or one of the non-GIST phenotypes, was confirmed by histology. The differential diagnosis for each patient is provided, where patients have label 1 if the lesion was a GIST, and label 0 if the lesion was a non-GIST. The subtype of the non-GIST lesions is also provided.

2.5. The CRLM dataset

This dataset consists of 77 patients with a total of 93 colorectal liver metastases (CRLM) with either a 100% desmoplastic histopathological growth patterns (HGP) [15] ($N = 46$) or 100% replacement HGP ($N = 47$), as described in Starmans et al. [13] [2]. For each patient, a portal venous phase CT scan is provided. The ground truth label, i.e., whether a lesion had a desmoplastic or replacement HGP, was determined on hematoxylin and eosin stained tissue sections. The HGP type for each patient is provided, where patients have label 1 if the lesions had replacement HGP, and label 0 if the lesions had a desmoplastic HGP. As the HGP is assumed to be the same for all lesions of a subject, the ground truth is provided on subject level.

For each patient, for each lesion, segmentations by three clinicians (STUD1, PhD, RAD) and a Convolutional Neural Network (CNN) are available: e.g. *"segmentation_lesion1_STUD1.nii.gz"*, *"segmentation_lesion1_PhD.nii.gz"*, *"segmentation_lesion1_RAD.nii.gz"*, and *"segmentation_lesion1_CNN.nii.gz"*. Additionally, each lesion was segmented a second time by the first observer (STUD2), and is named e.g. *"segmentation_lesion1_STUD2.nii.gz"*. Note that 8 out of the 93 lesions (9%) were missed by the CNN, and thus do not include a CNN segmentation

2.6. The Melanoma dataset

This dataset consists of 169 lung metastases of 103 patients with *BRAF* mutated ($N = 51$) or *BRAF* wild type ($N = 52$) metastatic melanoma, as described in Angus

[2] Starmans et al. [13] reported a total of 76 patients, but the dataset did actually contain 77 patients.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.19.21262238; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

et al. [14]. For each patient, a contrast-enhanced thoracic CT scan is provided. When multiple lesions were included, the corresponding segmentations are named *"segmentation_lesion1.nii.gz", "segmentation_lesion2.nii.gz",* and so on. The ground truth label, i.e., whether lesions from a patient were *BRAF* mutated or *BRAF* wild type, is provided, where patients have label 1 if the lesions were *BRAF* mutated, and label 0 if the lesions were *BRAF* wild type. As the *BRAF* mutation is assumed to be the same for all lesions of a subject, the ground truth is provided on subject level.

2.7. Acquisition protocol details

From the original DICOM files from the MRI and CT scans, the values of several tags were extracted to provide information on the used acquisition protocols, which for each scan are included in a *metadata.json* file.

For both MRI and CT scans, the following general acquisition protocol details from the following DICOM tags are included:

(0008, 0060) Modality
(0008, 0070) Manufacturer
(0008, 1090) Model name
(0018, 0020) Scanning sequence
(0018, 0022) Scan options
(0018, 0023) Acquisition type
(0018, 0024) Sequence name
(0018, 0050) Slice thickness
(0018, 0083) Number of averages
(0018, 0084) Spacing between slices
(0018, 0093) Percent sampling
(0018, 1030) Protocol name
(0018, 5100) Patient position
(0020, 0037) Orientation
(0028, 0030) Pixel spacing

For each MRI scan, the following specific acquisition protocol details from the following DICOM tags are additionally included:

(0018, 0080) Repetition time
(0018, 0081) Echo time
(0018, 0082) Inversion time
(0018, 0084) Imaging frequency
(0018, 0087) Tesla
(0018, 0091) Echo train length
(0018, 1250) Coil
(0018, 1310) Acquisition matrix
(0018, 1312) Encoding direction
(0018, 1314) Flip angle

medRxiv preprint doi: https://doi.org/10.1101/2021.08.19.21262238; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

For each CT scan, the following specific acquisition protocol details from the following DICOM tags are additionally included:

(0018, 0060) KVP (kilovoltage peak) **(0018, 1210)** Convolution kernel

3. Experimental Design, Materials and Methods

3.1. The Lipo dataset

Patients that were either referred to/discussed at, or diagnosed/treated at the Erasmus MC Cancer Institute, Rotterdam, the Netherlands, between December 2009 and August 2018 with a pathologically confirmed diagnosis of lipoma or WDLPS were retrospectively included. Inclusion criteria were: a known *MDM2* amplification status tested by fluorescence *in situ* hybridization (FISH); and at least a T1-weighted MRI sequence available before treatment (if applicable).

The lipoma and WDLPS lesions were segmented semi-automatically on the T1-weighted MRI [16]. All images were segmented independently by either a medical masters student or a PhD candidate with an MD degree. Both were blinded to the type of lipomatous lesion. To validate segmentation accuracy, a sample set was verified by a musculoskeletal radiologist, specialized in soft-tissue sarcomas (4 years of experience). Semi-automatic results were always reviewed and manually corrected when necessary, to assure the result resembled manual segmentation.

3.2. The Desmoid dataset

Patients that were either referred to/discussed at, or diagnosed/treated at the Erasmus MC Cancer Institute, Rotterdam, the Netherlands, between 1990 and 2018 with histologically proven primary or recurrent DTF, or a malignant extremity STS, were retrospectively included. Inclusion criteria were: at least a T1-weighted MRI sequence available before treatment (if applicable); for the STS, a histologically proven primary myxofibrosarcoma, myxoid liposarcoma or leiomyosarcoma of the extremities.

The DTF and STS were all segmented semi-automatically on the T1-weighted MRI [16]. All images were segmented independently by either a medical masters student or a PhD candidate with an MD degree under supervision of a musculoskeletal radiologist

medRxiv preprint doi: https://doi.org/10.1101/2021.08.19.21262238; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

(4 years of experience). Both were blinded to the type of lesion. Semi-automatic results were always reviewed and manually corrected when necessary, to assure the result resembled manual segmentation.

3.3. The Liver dataset

Patients that were either referred to/discussed at, or diagnosed/treated at the Erasmus MC Cancer Institute, Rotterdam, the Netherlands, between 2002 and 2018 with a primary solid liver lesion were retrospectively included. Inclusion criteria were: hepatocellular carcinoma (HCC), intrahepatic cholangiocarcinoma (iCCA), hepatocellular adenoma (HCA) or FNH; pathologically proven phenotype; and availability of a T2-weighted MRI scan. An exception to the pathologically proven phenotype was made for typical FNH, which are routinely not biopsied and diagnosed radiologically [17], as typical FNH imaging characteristics are 100% specific [18]. Exclusion criteria were: maximum diameter equal to or smaller than 3 cm; underlying liver disease; and significant imaging artefacts.

The lesions were all segmented semi-automatically on the T2-weighted MRI [16]. All images were segmented independently by one of two experienced abdominal radiologists (21 and 8 years of experience). Both were blinded to the type of lesion. Semi-automatic results were always reviewed and manually corrected when necessary, to assure the result resembled manual segmentation.

3.4. The GIST dataset

Patients that were either referred to/discussed at, or diagnosed/treated at the Erasmus MC Cancer Institute, Rotterdam, the Netherlands, between 2004 and 2017 with a histopathologically proven primary GIST or intra-abdominal tumors radiologically resembling GIST were retrospectively included. The inclusion criterion was availability of at least a contrast-enhanced venous-phase CT prior to treatment. The sample sizes of the non-GIST and the GIST cohort were matched. The non-GIST subtypes were balanced, i.e. a similar number of patients per subtype was randomly included.

The lesions were all segmented semi-automatically on the CT scan [16]. All images were segmented independently by either a medical masters student or a PhD candidate

medRxiv preprint doi: https://doi.org/10.1101/2021.08.19.21262238; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

with an MD degree under supervision of a musculoskeletal radiologist (5 years of experience). Both were blinded to the type of lesion. Semi-automatic results were always reviewed and manually corrected when necessary, to assure the result resembled manual segmentation.

3.5. The CRLM dataset

Patients that were surgically treated at the Erasmus MC Cancer Institute, Rotterdam, the Netherlands, between 2003 and 2015 with CRLM were included. Inclusion criteria were: availability of at least a contrast-enhanced venous-phase CT prior to treatment; available hematoxylin and eosin stained tissue sections; either a 100% desmoplastic HGP or a 100% replacement HGP. Exclusion criteria were: recurrent CRLM or CRLM requiring two-staged resections; and treatment with preoperative chemotherapy, since chemotherapy may alter HGPs [15]. HGPs were scored on resection specimens according to the consensus guidelines by an expert pathologist (PV) [19].

The lesions were all segmented semi-automatically on the CT scan [16]. Lesion segmentation was performed by four observers: a medicine student with no relevant experience (STUD1), a PhD student (PhD) with limited experience, an expert abdominal radiologist (RAD), and an automatic CNN. The student segmented all lesions a second time (STUD2). All observers were blinded to the type of lesion. Semi-automatic results were always reviewed and manually corrected when necessary, to assure the result resembled manual segmentation.

The CNN used for the automatic segmentations was the Hybrid-Dense-UNet, which achieved state-of-the-art performance on the LITS liver tumor segmentation challenge and is open-source [20, 21]. The original CNN as trained on the LITS data was used. From the CNN lesion segmentations, only lesions that had histology were extracted, and the segmentations were saved per lesion.

3.6. The Melanoma dataset

Patients that were diagnosed with metastatic melanoma at the Erasmus MC Cancer Institute, Rotterdam, the Netherlands, between January 2012 and February 2018 were

medRxiv preprint doi: https://doi.org/10.1101/2021.08.19.21262238; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

retrospectively included. Inclusion criteria were: known tumor *BRAF* mutation, diagnostic contrast-enhanced thoracic CT scan prior to commencement of any systemic therapy, and at least one lung metastasis of ≥ 10 mm evaluable according to Response Evaluation Criteria In Solid Tumors (RECIST) v1.1 [22]. Patients with *BRAF* mutations other than p.V600E were excluded. Formalin-fixed paraffin embedded material of the primary tumor and/ or metastasis was tested for *BRAF* (exon 15) using a polymerase chain reaction based assay or next generation sequencing as part of standard care.

Per patient, up to two lung lesions ≥ 10 mm were selected by a clinician supervised by an experienced chest radiologist and segmented semi-automatically on the CT scan [16]. In patients with >2 lung metastases of ≥10 mm, either the two largest or the two most easily distinguishable lesions were segmented (i.e., two separate lesions were preferred over two adjacent lesions). The clinician was blinded to the type of lesion. Semi-automatic results were always reviewed and manually corrected when necessary, to assure the result resembled manual segmentation.

4. Ethics Statement

The study protocol for the collection of the WORC database conformed to the ethical guidelines of the 1975 Declaration of Helsinki. Approval by the local institutional review board of the Erasmus MC (Rotterdam, the Netherlands) was obtained for collection of the WORC database (MEC-2020-0961), and separately for the six included studies (Lipo: MEC-2016-339, Desmoid: MEC-2016-339, Liver: MEC-2017-1035, GIST: MEC-2017-1187, CRLM: MEC-2017-479, Melanoma: MEC-2019-0693). The need for informed consent was waived due to the use of anonymized, retrospective data.

5. Acknowledgments

The authors thank Laurens Groenendijk for his assistance in processing the data and in the anonimization procedures. Martijn P. A. Starmans acknowledges funding from the research program STRaTeGy with project numbers 14929 and 14930, which

medRxiv preprint doi: https://doi.org/10.1101/2021.08.19.21262238; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

is (partly) financed by the Netherlands Organization for Scientific Research (NWO). Part of this study was financed by the Stichting Coolsingel (reference number 567), a Dutch non-profit foundation. This study is supported by EuCanImage (European Union's Horizon 2020 research and innovation programme under grant agreement Nr. 952103).

6. CRediT Author Statement

M.P.A.S., M.J.M.T., M.V., D.J.G., C.V., S.S., F.E.B., L.A., A.A.M.v.d.V., R.L.M., M.G., J.J.V., W.J.N., and S.K. provided the conception and design of the study. M.P.A.S., M.J.M.T., M.V., G.A.P., D.J.G., C.V., S.S., G.J.L.H.v.L., F.E.B., F.E.J.A.W., B.G.K., L.A., A.A.M.v.d.V., A.R., A.E.O., M.R., M.D., R.d.M., J.IJ., R.L.M., P.B.V., M.G.T., and J.J.V. acquired the data. M.P.A.S., M.J.M.T., M.V., F.E.B., L.A., R.L.M., M.G.T. and S.K. analyzed and interpreted the data. M.P.A.S. created the software. M.P.A.S. and S.K. drafted the article. All authors read and approved the final manuscript.

7. Declaration of Competing Interest

Wiro J. Niessen is founder, scientific lead, and shareholder of Quantib BV. Jacob J. Visser is a medical advisor at Contextflow. Astrid A. M. van der Veldt is a consultant (fees paid to the institute) at BMS, Merck, MSD, Sanofi, Eisai, Pfizer, Roche, Novartis, Pierre Fabre and Ipsen. The other authors do not declare any conflicts of interest.

References

[1] I. D. Nagtegaal, R. D. Odze, D. Klimstra, V. Paradis, M. Rugge, P. Schirmacher, K. M. Washington, F. Carneiro, I. A. Cree, the WHO Classification of Tumours Editorial Board, The 2019 WHO classification of tumours of the digestive system, Histopathology 76 (2020) 182–188. doi:10.1111/his.13975.

[2] J. Song, Y. Yin, H. Wang, Z. Chang, Z. Liu, L. Cui, A review of original articles published in the emerging field of radiomics, European Journal of Radiology 127 (2020) 108991. doi:10.1016/j.ejrad.2020.108991.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.19.21262238; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

[3] A. Traverso, L. Wee, A. Dekker, R. Gillies, Repeatability and reproducibility of radiomic features: A systematic review, International Journal of Radiation Oncology*Biology*Physics 102 (2018) 1143–1158. doi:10.1016/j.ijrobp.2018.05.053.

[4] M. L. Welch, C. McIntosh, B. Haibe-Kains, M. F. Milosevic, L. Wee, A. Dekker, S. H. Huang, T. G. Purdie, B. O'Sullivan, H. J. Aerts, D. A. Jaffray, Vulnerabilities of radiomic signature development: The need for safeguards, Radiotherapy and Oncology 130 (2019) 2–9. doi:10.1016/j.radonc.2018.10.027.

[5] D. S. Marcus, T. R. Olsen, M. Ramaratnam, R. L. Buckner, The extensible neuroimaging archive toolkit, Neuroinformatics 5 (2007) 11–33. doi:10.1385/ni:5:1:11.

[6] M. P. A. Starmans, S. R. Van der Voort, T. Phil, M. J. M. Timbergen, M. Vos, G. A. Padmos, W. Kessels, D. Hanff, D. J. Grünhagen, C. Verhoef, S. Sleijfer, M. J. van den Bent, M. Smits, R. S. Dwarkasing, C. J. Els, F. Fiduzi, G. J. L. H. Van Leenders, A. Blazevic, J. Hofland, T. Brabander, R. van Gils, G. J. H. Franssen, R. A. Feelders, W. W. De Herder, F. E. Buisman, F. E. J. A. Willemssen, B. Groot Koerkamp, L. Angus, A. A. M. Van der Veldt, A. Rajicic, A. E. Odink, M. Deen, J. M. Castillo T, J. F. Veenland, I. Schoots, M. Renckens, M. Doukas, R. A. De Man, J. N. M. Ijzermans, R. L. Miclea, P. B. Vermeulen, E. E. Bron, M. G. Thomeer, J. J. Visser, W. J. Niessen, S. Klein, Reproducible radiomics through automated machine learning validated on twelve clinical applications Submitted (2021).

[7] M. P. A. Starmans, 2021, WORCDatabase, Zenodo. URL: https://github.com/MStarmans91/WORCDatabase. doi:10.5281/zenodo.5119040.

[8] X. Li, P. S. Morgan, J. Ashburner, J. Smith, C. Rorden, The first step for neuroimaging data analysis: DICOM to NIfTI conversion, Journal of Neuroscience Methods 264 (2016) 47–56. doi:10.1016/j.jneumeth.2016.03.001.

[9] M. Vos, M. P. A. Starmans, M. J. M. Timbergen, S. R. van der Voort, G. A. Padmos, W. Kessels, W. J. Niessen, G. J. L. H. van Leenders, D. J. Grünhagen,

medRxiv preprint doi: https://doi.org/10.1101/2021.08.19.21262238; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

S. Sleijfer, C. Verhoef, S. Klein, J. J. Visser, Radiomics approach to distinguish between well differentiated liposarcomas and lipomas on MRI, British Journal of Surgery 106 (2019) 1800–1809. doi:10.1002/bjs.11410.

[10] M. J. M. Timbergen, M. P. A. Starmans, G. A. Padmos, D. J. Grünhagen, G. J. L. H. van Leenders, D. Hanff, C. Verhoef, W. J. Niessen, S. Sleijfer, S. Klein, J. J. Visser, Differential diagnosis and mutation stratification of desmoid-type fibromatosis on MRI using radiomics, European Journal of Radiology 131 (2020) 109266. doi:10.1016/j.ejrad.2020.109266.

[11] M. P. A. Starmans, R. L. Miclea, V. Vilgrain, M. Ronot, Y. Purcell, J. Verbeek, W. J. Niessen, J. N. M. Ijzermans, R. A. de Man, M. Doukas, S. Klein, M. G. Thomeer, Automated differentiation of malignant and benign primary solid liver lesions on MRI: an externally validated radiomics model, medRxiv (2021). doi:10.1101/2021.08.10.21261827.

[12] M. P. A. Starmans, M. J. M. Timbergen, M. Vos, M. Renckens, D. J. Grünhagen, G. J. L. H. van Leenders, R. S. Dwarkasing, F. E. J. A. Willemssen, W. J. Niessen, C. Verhoef, S. Sleijfer, J. J. Visser, S. Klein, Differential diagnosis and molecular stratification of gastrointestinal stromal tumors on CT images using a radiomics approach, arXiv:2010.06824 (2020).

[13] M. P. A. Starmans, F. E. Buisman, M. Renckens, F. E. J. A. Willemssen, S. R. Van der Voort, B. Groot Koerkamp, D. J. Grünhagen, W. J. Niessen, P. B. Vermeulen, C. Verhoef, J. J. Visser, S. Klein, Distinguishing pure histopathological growth patterns of colorectal liver metastases on CT using deep learning and radiomics: a pilot study, In Revision (2021).

[14] L. Angus, M. P. A. Starmans, A. Rajicic, A. E. Odink, M. Jalving, W. J. Niessen, J. J. Visser, S. Sleijfer, S. Klein, A. A. M. van der Veldt, The BRAF P.V600E mutation status of melanoma lung metastases cannot be discriminated on computed tomography by LIDC criteria nor radiomics using machine learning, Journal of Personalized Medicine 11 (2021) 257. doi:10.3390/jpm11040257.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.19.21262238; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

[15] S. Frentzas, E. Simoneau, V. L. Bridgeman, P. B. Vermeulen, S. Foo, E. Kostaras, M. R. Nathan, A. Wotherspoon, Z.-h. Gao, Y. Shi, G. V. den Eynden, F. Daley, C. Peckitt, X. Tan, A. Salman, A. Lazaris, P. Gazinska, T. J. Berg, Z. Eltahir, L. Ritsma, J. van Rheenen, A. Khashper, G. Brown, H. Nyström, M. Sund, S. V. Laere, E. Loyer, L. Dirix, D. Cunningham, P. Metrakos, A. R. Reynolds, Vessel co-option mediates resistance to anti-angiogenic therapy in liver metastases, Nature Medicine 22 (2016) 1294–1302. doi:10.1038/nm.4197.

[16] M. P. A. Starmans, S. Klein, S. R. van der Voort, M. G. Thomeer, R. L. Miclea, W. J. Niessen, Classification of malignant and benign liver tumors using a radiomics approach, in: E. D. Angelini, B. A. Landman (Eds.), Medical Imaging 2018: Image Processing, volume 10574, SPIE-Intl Soc Optical Eng, 2018, p. 105741D. doi:10.1117/12.2293609.

[17] V. Vilgrain, Focal nodular hyperplasia, European Journal of Radiology 58 (2006) 236–245. doi:10.1016/j.ejrad.2005.11.043.

[18] E. A. f. t. S. o. t. L. (EASL), EASL clinical practice guidelines on the management of benign liver tumours, Journal of Hepatology 65 (2016) 386–398. doi:10.1016/j.jhep.2016.04.001.

[19] P. B. Vermeulen, C. Colpaert, R. Salgado, R. Royers, H. Hellemans, E. V. den Heuvel, G. Goovaerts, L. Y. Dirix, E. V. Marck, Liver metastases from colorectal adenocarcinomas grow in three patterns with different angiogenesis and desmoplasia, The Journal of Pathology 195 (2001) 336–342. doi:10.1002/path.966.

[20] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, P.-A. Heng, H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes, IEEE Transactions on Medical Imaging 37 (2018) 2663–2674. doi:10.1109/tmi.2018.2845918.

[21] P. Bilic, P. F. Christ, E. Vorontsov, G. Chlebus, H. Chen, Q. Dou, C.-W. Fu, X. Han, P.-A. Heng, J. Hesser, S. Kadoury, T. K. Konopczynski, M. Le, C. Li, X. Li, J. Lipková, J. S. Lowengrub, H. Meine, J. H. Moltz, C. Pal, M. Piraud, X. Qi, J. Qi, M. Rempfler, K. Roth, A. Schenk, A. Sekuboyina, P. Zhou,

medRxiv preprint doi: https://doi.org/10.1101/2021.08.19.21262238; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

C. Hülsemeyer, M. Beetz, F. Ettlinger, F. Grün, G. Kaissis, F. Lohöfer, R. Braren, J. Holch, F. Hofmann, W. H. Sommer, V. Heinemann, C. Jacobs, G. E. H. Mamani, B. van Ginneken, G. Chartrand, A. Tang, M. Drozdzal, A. Ben-Cohen, E. Klang, M. M. Amitai, E. Konen, H. Greenspan, J. Moreau, A. Hostettler, L. Soler, R. Vivanti, A. Szeskin, N. Lev-Cohain, J. Sosna, L. Joskowicz, B. H. Menze, The liver tumor segmentation benchmark (LiTS), arxiv:1901.04056 (2019).

[22] E. Eisenhauer, P. Therasse, J. Bogaerts, L. Schwartz, D. Sargent, R. Ford, J. Dancey, S. Arbuck, S. Gwyther, M. Mooney, L. Rubinstein, L. Shankar, L. Dodd, R. Kaplan, D. Lacombe, J. Verweij, New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1), European Journal of Cancer 45 (2009) 228–247. doi:10.1016/j.ejca.2008.10.026.