Elsevier Editorial System(tm) for Radiotherapy and Oncology

Manuscript Draft

Manuscript Number: RO-D-18-00623R2

Title: CT-based radiomic features predict tumor grading and have prognostic value in patients with soft tissue sarcomas treated with neoadjuvant radiation therapy

Article Type: Full Length Article

Keywords: soft tissue sarcoma, radiomics, neoadjuvant radiotherapy, biomarker, machine learning, tumor grading, CT, imaging

Corresponding Author: Mr. Jan C Peeken, M.D.

Corresponding Author's Institution: Klinikum Rechts der Isar, Technical University of Munich

First Author: Jan C Peeken, M.D.

Order of Authors: Jan C Peeken, M.D.; Michael Bernhofer; Matthew B Spraker, MD; Daniela Pfeiffer, MD; Michal Devecka, MD; Ahmed Thamer, MD; Mohammed Shouman, MD; Armin Ott; Nüsslin Fridtjof, Prof, PhD; Nina A Mayr, Prof, MD; Burkhard Rost, Prof, PhD; Matthew J Nyflot, PhD; Stephanie E Combs, Prof, MD

Abstract: Purpose
In soft tissue sarcomas (STS) patients systemic progression and survival remain comparably low despite low local recurrence rates. In this work, we investigated whether quantitative imaging features ("radiomics") of radiotherapy planning CT-scans carry a prognostic value for pre-therapeutic risk assessment.

Methods
CT-scans, tumor grade, and clinical information were collected from three independent retrospective cohorts of 83 (TUM), 87 (UW) and 51 (McGill) STS patients, respectively. After manual segmentation and preprocessing, 1358 radiomic features were extracted. Feature reduction and machine learning modeling for the prediction of grading, overall survival (OS), distant (DPFS) and local (LPFS) progression free survival were performed followed by external validation.

Results
Radiomic models were able to differentiate grade 3 from non-grade 3 STS (area under the receiver operator characteristic curve (AUC): 0.64). The Radiomic models were able to predict OS (C-index: 0.73), DPFS (C-index: 0.68) and LPFS (C-index: 0.77) in the validation cohort. A combined clinical-radiomics model showed the best prediction for OS (C-index: 0.76). The radiomic scores were significantly associated in univariate and multivariate cox regression and allowed for significant risk stratification for all three endpoints.

Conclusion
This is the first report demonstrating a prognostic potential and tumor grading differentiation by CT-based radiomics.

CT-based radiomic features predict tumor grading and have prognostic value in patients with soft tissue sarcomas treated with neoadjuvant radiation therapy

Jan C. Peeken[1,2,3], Michael Bernhofer[4], Matthew B. Spraker[5], Daniela Pfeiffer[6], Michal Devecka[1], Ahmed Thamer[1], Mohammed A. Shouman[1], Armin Ott[7], Fridtjof Nüsslin[1,8], Nina A. Mayr[5], Burkhard Rost[4], Matthew J. Nyflot[5,9] & Stephanie E. Combs[1,2,3]

[1]Department of Radiation Oncology, Klinikum rechts der Isar, Technical University of Munich (TUM), Ismaninger Straße 22, 81675 Munich, Germany
[2]Institute of Innovative Radiotherapy (iRT), Department of Radiation Sciences (DRS), Helmholtz Zentrum München, Ingolstaedter Landstrasse 1, 85764 Neuherberg, Germany
[3]Deutsches Konsortium für Translationale Krebsforschung (DKTK), Partner Site Munich
[4]Department for Bioinformatics and Computational Biology, Informatik 12, Technical University of Munich (TUM), Boltzmannstraße 3, 85748, Garching, Germany
[5]University of Washington, Department of Radiation Oncology, Seattle, Washington, United States
[6]Department of Radiology, Klinikum rechts der Isar, Technical University of Munich (TUM), Ismaninger Straße 22, 81675 Munich, Germany
[7]Institut für Medizinische Statistik und Epidemiologie, Technical University of Munich (TUM), Ismaninger Straße 22, 81675 Munich, Germany
[8]Institute for Advanced Study (IAS), Technical University of Munich (TUM)
[9]University of Washington, Department of Radiology, Seattle, Washington, United States

Correspondence:
Jan C Peeken
email: jan.peeken@tum.de
tel: 00498941404501
Mailing address: Klinik für RadioOnkologie und Strahlentherapie, Universitätsklinikum rechts der Isar, Technische Universität München (TUM), Ismaninger Straße 22, 81675 München, Germany

Keywords: soft tissue sarcoma, radiomics, neoadjuvant radiotherapy, biomarker, machine learning, tumor grading, CT, imaging

Running head: CT-radiomics in soft tissue sarcomas

Funding:
This work was funded in part by research support for JP within the KKF program of the Medical Faculty of the Technical University of Munich (TUM), Deutsches Konsortium für Translationale Krebsforschung (DKTK), Partner Site Munich (to JP, SC), German Research Foundation (DFG), and the Bavarian Competence Network for Technical and Scientific High-Performance Computing (to MB and BR). MJN received support from the R&E Foundation of the Radiological Society of North America during the study.

Number of Figures: 3
Number of Tables: 3
Number of Text Pages: 15
Supplemental Materials: 3
Number of Supplemental Figures: 7
Number of Supplemental Tables: 5

Abstract

Purpose

In soft tissue sarcomas (STS) patients systemic progression and survival remain comparably low despite low local recurrence rates. In this work, we investigated whether quantitative imaging features ("radiomics") of radiotherapy planning CT-scans carry a prognostic value for pre-therapeutic risk assessment.

Methods

CT-scans, tumor grade, and clinical information were collected from three independent retrospective cohorts of 83 (TUM), 87 (UW) and 51 (McGill) STS patients, respectively. After manual segmentation and preprocessing, 1358 radiomic features were extracted. Feature reduction and machine learning modeling for the prediction of grading, overall survival (OS), distant (DPFS) and local (LPFS) progression free survival were performed followed by external validation.

Results

Radiomic models were able to differentiate grade 3 from non-grade 3 STS (area under the receiver operator characteristic curve (AUC): 0.64). The Radiomic models were able to predict OS (C-index: 0.73), DPFS (C-index: 0.68) and LPFS (C-index: 0.77) in the validation cohort. A combined clinical-radiomics model showed the best prediction for OS (C-index: 0.76). The radiomic scores were significantly associated in univariate and multivariate cox regression and allowed for significant risk stratification for all three endpoints.

Conclusion

This is the first report demonstrating a prognostic potential and tumor grading differentiation by CT-based radiomics.

Introduction

Soft tissue sarcomas (STS) constitute an overall rare malignant entity comprising 1% of all cancers [1]. Treatment and outcome differ vastly between anatomic sites of occurrence [1,2]. Resection of high risk STS of the extremity is combined with neoadjuvant or adjuvant radiotherapy (RT) improving locoregional control (LC) and eventually survival [3,4]. Compared to adjuvant RT, neoadjuvant RT offers several advantages including lower total radiation doses, smaller target volumes, and reduced late toxicities, such as fibrosis[5,6]. There is also some evidence showing improved survival in favor to neoadjuvant RT [7,8]. In contrast to the high LC rates of up to 94%, current therapy regiments achieve comparably low systemic control rates and overall survival of 61% and 64%, respectively [9].

There is an ongoing search for biomarkers as prognostic factors. Despite, large research efforts STS histologies or molecular aberrations have not yet been established as prognostic markers. Therapy decisions are still mostly made using basic clinical determinants such as TNM staging and grading.

Imaging-based radiomics constitutes an alternative tool to characterize tissue. It is defined as an algorithm-based large-scale quantitative analysis of imaging features [10]. Radiomics has been associated with survival, tumor progression, and molecular changes including genetic mutations or expression profiles as shown in multiple malignant entities [11–14]. In contrast to pathology, radiomics has the principal advantage of analyzing the whole tumor before therapy rather than a focal biopsy.

In this study we analyzed the prognostic capability of pre-RT planning CT-based radiomics machine learning (ML)-based prognostic models. A radiomic model was generated to predict French Fédération Nationale des Centres de Lutte Contre le Cancer (FNCLCC) grade as a biological correlate. The resulting models were tested on external validation cohorts.

Methods

Patients

Radiomic model training and validation were performed with three independent patient cohorts treated for STS at the Technical University of Munich (TUM), the University of Washington/Seattle Cancer Care Alliance (UW) and McGill University. Clinical information including age, TNM-staging and grading were assessed for all patients. All patients received neoadjuvant RT followed by tumor resection with or without chemotherapy. No patient received previous RT. The dataset from McGill University (McGill) published by Vallières et al. in the "The Cancer Imaging Archive" was added as additional validation and training set [15,16]. Exclusion criteria for all three datasets were osteosarcomas, Ewing sarcomas, endoprothesis-dependent CT artifacts, incomplete imaging and direct affection of bones (see Supplemental Figure 1 for patient workflow).

The overall survival (OS) was calculated from initial pathologic diagnosis to the time point of death or the time point of censoring. Distant progression free survival (DFPS) and local progression free survival (LPFS) were calculated from diagnosis to the first sign of progression, death, or time point of censoring. Approval from the ethic committees was received from both institution. Informed consent was given before therapy.

Image acquisition and definition of volume of interests

In the TUM and UW cohorts, each patient received a planning CT before RT as published previously [17]. In the McGill cohort, CT data were obtained from pre-therapeutic PET/CT studies. See Supplemental Table 2 for acquisition parameters. Contrast agent was administered at TUM in 8 patients (9.6%) (70 ml Imeron 400 MCT, Bracco Imaging, Germany), in 77 patients at UW (88.5%) (100ml Omnipaque 300, GE Healthcare, USA) and 0 patients in the McGill data set (see supplemental data for imaging protocols).

For segmentation, iplan RT 4.1.2 (Brainlab, Munich, Germany) and Eclipse 13.0 (Varian Medical Systems, Palo Alto, USA) were used at TUM whereas MIM software (version 6.6, MIM Software Inc, Cleveland, USA) were used at UW. As volume of interest (VOI) the primary tumor was manually segmented by a radiation oncology

resident (JP), by adapting existing expert segmentations from RT treatment planning in the following ways: MRI-scans were considered if available, edematous changes were not included into the VOI, and bone structures were subtracted from the VOI.

To reduce operator-dependent bias, multiple segmentations were performed for 20 randomly selected patients in the TUM cohort by three radiation oncologists (JP, TA, MS) (see Figure 1). Dice coefficients (DC) were calculated by using the DiceComputation module of 3D Slicer (3D Slicer, Version 4.8 stable release) [18].

Image preprocessing and radiomic feature extraction

All preprocessing steps and radiomic feature extraction were performed using the pyradiomics package (Version 1.3) in Python (3.6.4) [19]. Preprocessing included image discretization with a fixed bin width of 10 (Supplemental Material 1 for a closer definition).

Isotropic resampling was performed by Bspline interpolation to receive a voxel size of 1x1x1 mm. Preprocessing included image discretization with a fixed bin width of 10 (see Supplemental Material 1 for the equation). Besides the original images, image reconstructions were performed using Laplacian of Gaussian filtering (Sigma values 1.0, 2.0, 3.0, 4.0, and 6.0) and wavelet filtering yielding 8 decompositions. Besides 14 shape features, 1344 were extracted from original and filtered reconstructions of the images including intensity histogram (first-order) and texture features. Texture parameters included "Gray Level Co-occurrence Matrix" (GLCM) features, "Gray Level Size Zone Matrix" (GLSZM) features, "Gray Level Run Length Matrix" (GLRLM) features, "Neighbouring Gray Tone Difference Matrix" (NGTDM) features, and "Gray Level Dependence Matrix" (GLDM) features. See Supplemental Material 1 for a detailed listing of extracted features.

Radiomics feature reduction

We applied two steps of unsupervised feature reduction prior to training our models. First, we calculated the intraclass correlation coefficient ICC(3,1) for all 1358 radiomics features based on the above-mentioned resegmentations (20 patients with three observations each) for the TUM data set in order to estimate how small differences in the image annotation influenced the selected features. We kept features with ICC(3,1) ≥ 0.8. Next, we removed inter-correlated features: for each

pair of features *F1* and *F2* we removed *F2* if the values of both features had an absolute Pearson correlation coefficient of 0.9 or higher across all patients. Additionally, the following six clinical features were used as clinical input features: age, FNSSC tumor grading, T-Suffix, T-stage, M-stage, and N-stage.

During training we standardized each feature by subtracting its median from its values and dividing the results by its interquartile range (IQR). This method is usually more robust against outliers than the default standardization (subtracting the mean and dividing by the standard deviation). Further, we tested supervised feature selection by ranking the features according to their predictive power and keeping only the best ten. The predictive power of a feature was estimated by removing all other features and using this feature alone to predict.

Except for the two unsupervised feature set reductions, all other steps were performed during a five-fold cross-validation to prevent bias (cf. model training and validation).

Machine learning model design

We designed four different models for each of the three clinical endpoints: OS, DPFS and LPFS. Each model was based on gradient boosting with component-wise least squares as base learners designed for survival analysis [20].

The *Radiomics* model was trained on all radiomic features while the *Clinical* model used only the clinical features. In addition, there were two combination models: one was trained on the clinical features and the tumor volume (*Combined-Volume* model), the other used all radiomics and clinical features (*Combined-Radiomics*).

We also trained two random forest models to predict FNSSC tumor grading: one using only the radiomics features, the other using only the clinical features. Due to the lack of patients with grading G1 in both, the TUM and McGill data set, we classified binary: grading G3 or not, i.e. G1 or G2.

All models were built using Python (3.6.4) and the machine learning libraries scikit-learn (0.19.2) and scikit-survival (0.6.0) [21–24]. A documentation can be found in the Supplemental Material 1. The models are provided online. Selected radiomic features and respective indices are provided in Supplemental Material 2.

Model training and validation

In order to minimize bias and establish model validity on the TUM data set, we trained and tested all models within a nested cross-validation: five outer folds with five inner folds each. Feature selection and hyperparameter optimization via grid-search was performed as part of the inner cross-validation. Prediction performance in the form of the concordance index (CI) or area under the receiver operating characteristic curve (AUC) was established on the outer cross-validation and averaged over the five folds. The 95-percent confidence intervals were estimated via bootstrapping (1000 iterations).

The final models were trained on the whole TUM data set using the feature sets and parameters with the best average performance over all 25 inner cross-validation splits. Final performance was tested on the UW data set to get an unbiased estimate.

The *Radiomics* model was also tested on the McGill cohort. In addition, we retrained the *Radiomics* model on the combined TUM and McGill data sets (*Radiomics - Retrained* model) to test the effect of more training samples (see Supplemental Figure 1(B) for patient workflow and Supplemental Table 3 for C-indices). Clinical parameters were not available for the McGill cohort for calculation of the other models.

Comparison to single-predictor models

We also compared the performance of our machine learning models for the clinical endpoints to simple single-predictor models. The *Volume* model used only the tumor volume, the *AJCC* model only the TNM staging system based on the 7th edition of the American joint committee on Cancer (AJCC) [25].

Difficult feature reduction

The first feature reduction based on the ICC(3,1) threshold of at least 0.8 resulted in a set of 623 stable features (Supplemental Material 2). Removing all highly inter-correlated features (R >= 0.9) yielded a final set of 127 non-correlated features.

The supervised feature selection, i.e. keeping only the 10 most predictive features, resulted in rather different feature sets for each of the cross-validation folds. Due to this unstable behavior, most likely caused by the small data set size, we ultimately decided against it and trained the models on all 127 non-correlated features. Additionally, the random forest models for the grading prediction exhibited better

performance when trained on all 623 stable features, instead. See Supplemental Figures 2-3 for distribution of intensity and volume features.

Influence of contrast-enhancement on feature values

In order to assess the dependency of the selected features on the usage of contrast agent, we compared feature values between the patients that either received contrast agent or not. Of all non-correlated features, 28 were significantly different between both groups, however, no feature retained significance after adjustment for multiple testing (Supplemental Table 5).

Statistics

All statistical analyses were performed in R (version 3.4.0) (R core team, Vienna, Austria). Classified patient subgroups were analyzed with Kaplan-Meier survival curves using the "ggkm" package. Dichotomization was performed using the median of the respective predictor as cut-point. The log-rank test was used to assess statistical significance. Univariate and multivariate Cox proportional hazard models were used to assess prognostic significance of features using the "survival" package. Receiver operator characteristic curves (ROC) and respective area under the curve (AUC) values were calculated using the "survivalROC" package. Calibration curves were generated using the packages "rms" and "riskRegression". Bonferroni correction was performed in cases of multiple testing. A p-value below 0.05 was regarded as significant.

Results

The flowchart of the study is depicted in Figure 1.

Patient characteristics and VOI definition

Patient demographics, RT schedules, and histologies were similar between groups (see Table 1 and Supplemental Table 1). Chemotherapy was significantly more frequently given in the UW patient cohort ($p<0.0001$). Chemotherapy appeared to be a significant prognostic factor for OS ($p=0.003$) and DPFS ($p=0.005$) in the TUM cohort as well as for DPFS in the McGill cohort ($p=0.0016$). In contrast to TUM and UW, 10% of patients in the McGill cohort did not receive RT ($p<0.0001$). The TUM cohort had significantly more adverse features including 9 recurrent tumors ($p=0.002$) and higher AJCC stages ($p<0.0001$) (e.g. stage III TUM: 75.9%, UW: 43.7%). Median OS, DPFS and LPFS were non-significantly longer in the UW cohort. Similarity

between multiple delineation was overall high with a mean DC over all three operators of 0.91 (standard deviation 0.069).

Training and validation of prognostic radiomic classifiers

For the prediction task of OS, the *Radiomic* model achieved a predictive performance with a C-index of 0.72 (0.61-0.83) on the training set (see Supplemental Figure 1(A) for patient workflow, Supplemental Figures 4-6 for calibration curves, Figure 2 for ROC curves and Table 2 for C-indices). This performance was about 14% better than the *AJCC* staging system (C-index: 0.63 (0.56-0.60)) but about 5% worse than the *Clinical* model (C-index: 0.76 (0.68-0.81). Both differences marked tendencies that were statistically insignificant. Validation on the UW patient set successfully reproduced a stable performance of the *Radiomics* model with a C-index of 0.73 (0.63-0.82). The *Clinical* model (C-index: 0.69 (0.57 - 0.80)) and the *AJCC* staging system (C-index: 0.68 (0.57 - 0.77)) performed worse in the validation set. Validation on the McGill cohort yielded a C-index of 0.59 (0.39-0.79).

Radiomics-based prediction of DPFS yielded overall lower predictive performances with a C-index of 0.64 (0.47-0.70) in the training set. This was significantly better compared to the *Clinical* model (C-index: 0.46 (0.42-0.53)), although not significantly different from random. In the validation cohorts, predictive performances were similar among all groups (*Radiomics*: UW C-index: 0.68 (0.55-0.76), McGill C-index: 0.73 (0.61-0.83); *Clinical*: C-index: 0.66 (0.55-0.77), AJCC: 0.66 (0.54-0.74)).

For the prediction of LPFS the *Clinical* model (C-index: 0.68 (0.51-0.81) achieved a higher prediction compared the *Radiomics* model (C-index: 0.56 (0.40-0.71)) and the *AJCC* staging system (C-index: 0.57 (0.51-0.63)). In the validation set however, the *Radiomics* model significantly outperformed the *Clinical* model and the *AJCC* staging system with a C-index of 0.77 (0.66-0.87). Due to only one event in the McGill cohort, no meaningful outcome measure was calculated.

Furthermore, we tested the prognostic power of tumor volume alone. Predicting OS and DPFS, *Volume* performed non-significantly worse compared to the radiomics model in the validation set (differences of C-index of -0.08 and -0.03, respectively). For LPFS, *Volume* did not show a significant predictive value.

To asses the incremental value of radiomics, joint models combining the radiomic features or volume with clinical features were developed. The *Combined-Radiomics* model performed similar to the *Clinical* model in the training set (C-index: 0.75 (0.70-0.80)) but non-significantly worse compared to the *Combined-Volume* model (C-index: 0.78 (0.74-0.84) predicting OS on the training set. However, both combined models significantly outperformed the *AJCC* staging system. In the test set (UW), the *Combined-Radiomics* model retained its predictive performance (C-index 0.76 (0.68-0.81) in contrast to the *Clinical* and *Combined-Volume* model (C-index 0.72 (0.60-0.83)). For DPFS prediction, both combined models performed similar to the *Radiomics* model. Predictive performance for LPFS was reduced for the *Combined-Radiomics* model compared to *Radiomics* alone. *Combined-Volume* did not predict LPFS significantly better than random.

The *Radiomics - Retrained* model achieved overall lower predictive performances on the combined training set (see Supplemental Table 3 for C-indices). The resulting models achieved significant predictions for all three prediction tasks on the validation set with similar performances as described above.

The radiomics score as a prognostic factor

Univariate cox regression analysis in the training cohort identified the *Radiomics* score, *Volume*, age and M-stage as significant prognostic factors for OS (see Table 3 for included variables hazard ratios and p-values). For DPFS and LPFS, the *Radiomics* scores were the only significant factors. In multivariate analysis, the *Radiomics* score for DPFS was the only significant predictors. In the validation cohort, the *Radiomics* scores were significantly associated with all three endpoints in univariate and multivariate analyses. Besides, *Volume* and grading were significantly associated with OS and DPFS. However, only grading retained significance in the multivariate analysis for DPFS. There was a significant association of the *Radiomics* score for DPFS in the McGill set.

Patient risk stratification

Radiomics model-based risk stratification of patients did show a significant separation of survival curves for OS (p=0.00214) and DPFS (p=.0277) in the TUM cohort (see Figure 3). On the UW cohort the models yielded significant risk stratification with separation of survival curves for patients' OS (p=0.0015), DPFS (p=0.0024) and LPFS (p=0.00268).

Feature importance

The gradient boost predictor assigned feature coefficients during training. Dependent on the endpoint, 13 to 35 features were assigned with non-zero coefficients (see Supplemental Table 5). "GLRL_LowGrayLevelRunEmphasis" and "Firstorder_Skewness" were rated among the top four features for all prediction tasks. "Sphericity" appeared to be the most important feature for OS and LPFS prediction.

A radiomic classifier significantly differentiates tumor grading

The *Radiomics* classifier for tumor grading achieved a significant classification performance in the combined training cohort with an AUC of 0.65 (0.54 - 0.75). The classifier was successfully validated in the UW cohort with an AUC of 0.64 (0.52 - 0.76). For comparison, a *Clinical* model trained on the TUM set did not predict grading significantly different from random (TUM-training: AUC 0.62 (0.48-0.75), UW-testing: AUC 0.51 (0.38-0.63). See Supplemental Material 2 for feature importance and Supplemental Figure 7 for calibration curves.

Discussion

This is the first study to show a prognostic value for CT-based radiomic features in STS despite the low soft tissue contrast of CT. Our radiomic models showed predictive performances for patients' OS, DPFS and LPFS. Results were successfully reproduced in an external validation cohort. The radiomic scores were significantly associated with all three endpoints in univariate and multivariate cox regression and allowed for significant risk stratification for OS, DPFS and LPFS in the validation cohort. In comparison to a *Clinical* model, the *Radiomics* model showed more consistent results and a significantly better LPFS prediction on the validation set. Combining the clinical and radiomic features achieved the best overall performance for OS, however without a significant difference.

Our radiomic grading model differentiated grade 3 from non-grade 3 tumors significantly better than random in both training and validation sets. With a maximum AUC of 0.65, however, the discriminative capacity would not be sufficient for a potential substitution of invasive biopsies and pathological work-up.

A recent study indicated that performances of radiomic models may often be caused by radiomic feature correlation to tumor volume [26]. We tried to eliminate volume-correlated features by using feature inter-correlation for feature reduction. In direct comparison on the validation set, C-indices of *Volume* were non-significantly lower than the *Radiomics* model for OS or DPFS and significantly lower for LPFS. The *Combined-Volume* models also showed a non-significant lower performance for OS, significantly lower prediction for LPFS, but a non-significantly better performance for DPFS on the validation set compared to the *Combined-Radiomics* model. In addition, Volume did not retain significance in multivariate cox regression. These observations may indicate an incremental benefit of Radiomics above tumor volume alone for prediction of LPFS and potentially for OS.

The AJCC staging system performed overall worse on both patient cohorts compared to the *Clinical* and *Radiomics* model. Over all endpoints it performed better on the UW cohort compared to the TUM cohort. The difference in prognostic performance could be explained by the composition of histological STS subtypes between both cohorts. In addition, AJCC stages were more widely distributed among patients of the UW cohort whereas the TUM cohort was dominated by patients of stage III (75% of patients).

Regarding independent validation, the *Combined-Radiomics* model achieved the highest performances for prediction of OS, whereas the *Radiomics model* constituted the best performing model for LPFS. Such improved patient risk stratification may be utilized for personalized treatment. Current therapy regiments include resection and radiotherapy yielding good LC but low systemic control rates. Multiple studies analyzed the potential use of chemotherapy at different time points. For instance, a prospective trial administered a chemotherapy regiment including mesna, adriamycin, ifosfamide and dacarbazine (MAID) in addition to surgery and RT. Long term results showed an excellent 7y disease specific survival of 71%, but with the high toxicity rates [27]. Novel approaches, such as targeted therapies (e.g. angiogenesis or cell cycle inhibitors) would constitute further options for a systemic therapy escalation [28,29]. RT escalations by delivery of simultaneous integrated boosts may constitute a further alternative for therapy intensification [30]. High-risk

patients identified by a radiomic model could be administered to such additional systemic therapies whereas low-risk patients could be spared unnecessary toxicities.

Immunotherapy may constitute an alternative novel treatment option. However, early data suggested that immunotherapy may not be effective in unselected STS patients [31–33]. Selected STS subgroups such as undifferentiated pleomorphic sarcoma (UPS) or dedifferentiated liposarcomas showed higher response rates [31]. In coherence, UPS were recently shown to express higher levels of programmed cell death protein (PD-1) and programmed death-ligand 1 (PD-L1) as well as higher T-cell infiltration [31,34]. Radiomics directed to predict the immunogenic phenotype may provide in alternative tool to define immune-responsive STS independent of histologies.

Magnetic resonance imaging (MRI) allows for detailed imaging of soft tissues, making it the primary imaging modality in the management of STS and initial studies have explored the potential of MRI-based radiomics. Vallières et al. built a prognostic model by combining four textural features from fused FDG-PET/T1-weight and FDG-PET/ T2-weight scans for the prediction of lung metastases in the McGill cohort used in this study [15]. Recently, Spraker et al demonstrated prognostic relevance of T1-weight sequence-based radiomics parameters which was successfully validated on our patient cohort [35]. Spatial heterogeneity of FDG-PET uptake was also shown to be an independent prognostic predictor [36]. In this work we demonstrated the feasibility of radiomic-based prognostic risk assessment despite the inferior soft tissue resolution of CT. Further studies should evaluate a potential incremental benefit by combining radiomics features from CT, PET and MRI imaging.

There are several limitations of the study. First of all, the retrospective nature of this study may be a reason for potential bias [37]. Secondly, available patient numbers were comparably low, especially in the patient training set. Consequently, large standard deviations of performance metrics made direct comparison between models difficult and impeded interpretability of multivariate models. Thirdly, the large technical variances in image acquisition, such as CT scanner type or the inverse rate of contrast agent usage between cohorts, may have affected prediction performances. Radiomic features are known to be sensitive to aspects of image

acquisition and reconstruction. The reduced predictive performance after retraining the ML models on the combined TUM and McGill cohort may have been caused by the variability between the three distinct CT scanners and imaging protocols (two types of planning CTs and one hybrid PET/CT). However, the retention of significance in the validation dataset, which used different scanner hardware and protocols, may suggest at least some of the investigated image features are robust to variability for sarcoma CT imaging [38,39]. Generally, a future sufficiently sized prospective trial may overcome these limitations, however the data provide a valuable basis for further research in this field and underline the potential of radiomics and biomarkers in radiation oncology.

Calibration of predictive models for OS and PFS generally showed good calibration in concordance with the respective discriminate capacities. For LPFS, however, the *Radiomics* model showed suboptimal calibration despite a good discrimination with a C-index of 0.77 in the validations et (Supplementary Figure 4F). The stability of the model may have been affected by the relatively low event number (n=15). Interestingly, by adding clinical information to the *Combined Radiomics* model an improved calibration was achieved (Supplemental Figure 6F).

To conclude, for the first time we could demonstrate the prognostic potential of radiotherapy planning CT-based radiomic models. *Radiomics* models allowed significant patient stratification for OS, DPFS and LPFS. External validation was successful despite large technical variances.

Acknowledgments

We sincerely thank Dr Tobias Chapman, Dr Meghan Macomber and Dr. Kevin Ball for expert segmentations.

Conflict of interest statement:

The authors declare no potential conflicts of interest.

References

[1] Gutierrez JC, Perez EA, Franceschi D, et al. Outcomes for Soft-Tissue Sarcoma in 8249 Cases from a Large State Cancer Registry. J Surg Res 2007;141:105–14. doi:10.1016/j.jss.2007.02.026.

[2] Spraker MB, Bair E, Bair R, et al. An analysis of patient characteristics and clinical outcomes in primary pulmonary sarcoma. J Thorac Oncol 2013;8:147–51. doi:10.1097/JTO.0b013e318277401f.

[3] Gerrand CH, Rankin K. The treatment of soft-tissue sarcomas of the extremities. Prospective randomized evaluations of (1) limb-sparing surgery plus radiation therapy compared with amputation and (2) the role of adjuvant chemotherapy. Class Pap Orthop 2014:483–4. doi:10.1007/978-1-4471-5451-8_125.

[4] Koshy M, Rich S, Mohiuddin M. Improved Survival with Radiation Therapy in High Grade Soft Tissue Sarcomas of the Extremities: A SEER Analysis. Int J Radiat Oncol Biol Phys 2013;77:1–15. doi:10.1016/j.ijrobp.2009.04.051.Improved.

[5] O'Sullivan B, Davis AM, Turcotte R, et al. Preoperative verses postoperative radiotherapy in soft-tissue sarcoma of the limbs: a randomised trial. Lancet 2002;359:2235–41. doi:10.1016/S0140-6736(02)09292-9.

[6] Lartigau E, Kantor G, Taieb S, et al. [Definitions of target volumes in soft tissue sarcomas of the extremities]. Cancer Radiother 2001;5:695–703.

[7] Al-Absi E, Farrokhyar F, Sharma R, et al. A systematic review and meta-analysis of oncologic outcomes of pre- versus postoperative radiation in localized resectable soft-tissue sarcoma. Ann Surg Oncol 2010;17:1367–74. doi:10.1245/s10434-009-0885-7.

[8] Sampath S, Schultheiss TE, Hitchcock YJ, et al. Preoperative versus

postoperative radiotherapy in soft-tissue sarcoma: Multi-institutional analysis of 821 patients. Int J Radiat Oncol Biol Phys 2011;81:498–505. doi:10.1016/j.ijrobp.2010.06.034.

[9] Alektiar KM, Brennan MF, Healey JH, Singer S. Impact of intensity-modulated radiation therapy on local control in primary soft-tissue sarcoma of the extremity. J Clin Oncol 2008;26:3440–4. doi:10.1200/JCO.2008.16.6249.

[10] Peeken JC, Nüsslin F, Combs SE. “Radio-oncomics” - The potential of radiomics in radiation oncology. Strahlenther Onkol 2017;193:767–79. doi:10.1007/s00066-017-1175-0.

[11] Aerts HJWL, Velazquez ER, Leijenaar RTH, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5:4006. doi:10.1038/ncomms5006.

[12] Rios Velazquez E, Parmar C, Liu Y, et al. Somatic mutations drive distinct imaging phenotypes in lung cancer. Cancer Res 2017;77:3922–30. doi:10.1158/0008-5472.CAN-17-0122.

[13] Mattonen SA, Palma DA, Johnson C, et al. Detection of Local Cancer Recurrence After Stereotactic Ablative Radiation Therapy for Lung Cancer: Physician Performance Versus Radiomic Assessment. Int J Radiat Oncol 2016;94:1121–8. doi:10.1016/j.ijrobp.2015.12.369.

[14] Panth KM, Leijenaar RTH, Carvalho S, et al. Is there a causal relationship between genetic changes and radiomics-based image features? An in vivo preclinical experiment with doxycycline inducible GADD34 tumor cells. Radiother Oncol 2015;116:462–6. doi:10.1016/j.radonc.2015.06.013.

[15] Vallières M, Freeman CR, Skamene SR, El Naqa I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. Phys Med Biol 2015;60:5471–96.

doi:10.1088/0031-9155/60/14/5471.

[16] Clark K, Vendt B, Smith K, et al. The cancer imaging archive (TCIA): Maintaining and operating a public information repository. J Digit Imaging 2013;26:1045–57. doi:10.1007/s10278-013-9622-7.

[17] Röper B, Heinrich C, Kehl V, et al. Study of Preoperative Radiotherapy for Sarcomas of the Extremities with Intensity-Modulation, Image-Guidance and Small Safety-margins (PREMISS). BMC Cancer 2015;15:904. doi:10.1186/s12885-015-1633-y.

[18] Fedorov A, Beichel R, Kalphaty-Cramer J, et al. 3D slicers as an image computing platform for thw quantitative imaging network. Magn Reson Imaging 2012;30:1323–41. doi:10.1016/j.mri.2012.05.001.3D.

[19] van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Res 2017;77:e104–7. doi:10.1158/0008-5472.CAN-17-0339.

[20] Hothorn T, Bühlmann P, Dudoit S, Molinaro A, Van Der Laan MJ. Survival ensembles. Biostatistics 2006;7:355–73. doi:10.1093/biostatistics/kxj011.

[21] Pölsterl S, Navab N, Katouzian A. Fast Training of Support Vector Machines for Survival Analysis. Effic. Learn. Mach., vol. 1, Berkeley, CA: Apress; 2015, p. 243–59. doi:10.1007/978-3-319-23525-7_15.

[22] Pölsterl S, Navab N, Katouzian A. An Efficient Training Algorithm for Kernel Survival Support Vector Machines 2016;1. doi:10.1007/978-3-319-23525-7_15.

[23] Pölsterl S, Gupta P, Wang L, et al. Heterogeneous ensembles for predicting survival of metastatic , castrate-resistant prostate cancer patients [version 1 ; referees : 3 approved with reservations] Referee Status : 2018:1–28.

[24] Pedregosa F, Weiss R, Brucher M. Scikit-learn Machine Learning in Python 2011;12:2825–30. doi:10.1007/s13398-014-0173-7.2.

[25] Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th Edition of the AJCC Cancer Staging Manual and the Future of TNM. Ann Surg Oncol 2010;17:1471–4. doi:10.1245/s10434-010-0985-4.

[26] Welch ML, Mcintosh C, Haibe-Kains B, et al. Vulnerabilities of radiomic signature development: The need for safeguards. Radiother Oncol 2018. doi:10.1016/j.radonc.2018.10.027.

[27] Mullen JT, Kobayashi W, Wang JJ, et al. Long-term follow-up of patients treated with neoadjuvant chemotherapy and radiotherapy for large, extremity soft tissue sarcomas. Cancer 2012;118:3758–65. doi:10.1002/cncr.26696.

[28] Wong P, Houghton P, Kirsch DG, et al. Combining targeted agents with modern radiotherapy in soft tissue sarcomas. J Natl Cancer Inst 2014;106:16–8. doi:10.1093/jnci/dju329.

[29] Schwartz GK, Tap WD, Qin L-X, et al. Cixutumumab and temsirolimus for patients with bone and soft-tissue sarcoma: a multicentre, open-label, phase 2 trial. Lancet Oncol 2013;14:371–82. doi:10.1016/S1470-2045(13)70049-4.

[30] DeLaney TF, Chen YL, Baldini EH, et al. Phase 1 trial of preoperative image guided intensity modulated proton radiation therapy with simultaneously integrated boost to the high risk margin for retroperitoneal sarcomas. Adv Radiat Oncol 2017;2:85–93. doi:10.1016/j.adro.2016.12.003.

[31] Tawbi HA-H, Burgess MA, Crowley J, et al. Safety and efficacy of PD-1 blockade using pembrolizumab in patients with advanced soft tissue (STS) and bone sarcomas (BS): Results of SARC028—A multicenter phase II study. J Clin Oncol 2016;34:Supple 11006-11006.

[32] Pollack SM, Ingham M, Spraker MB, Schwartz GK. Emerging targeted and immune-based therapies in sarcoma. J Clin Oncol 2018;36:125–35. doi:10.1200/JCO.2017.75.1610.

[33] Wisdom AJ, Mowery YM. Rationale and Emerging Strategies for Immune Checkpoint Blockade in Soft Tissue Sarcoma 2018:1–11. doi:10.1002/cncr.31517.

[34] Pollack SM, He Q, Yearley JH, et al. T-cell infiltration and clonality correlate with programmed cell death protein 1 and programmed death-ligand 1 expression in patients with soft tissue sarcomas. Cancer 2017;123:3291–304. doi:10.1002/cncr.30726.

[35] Spraker MB, Wootton L, Hippe DS, et al. Radiomic Signature Extracted from Magnetic Resonance Imaging Predicts Outcomes in Soft Tissue Sarcoma. Int J Radiat Oncol Biol Phys 2017;99:S78–9. doi:10.1016/j.ijrobp.2017.06.190.

[36] Eary JF, O'Sullivan F, O'Sullivan J, Conrad EU. Spatial Heterogeneity in Sarcoma 18F-FDG Uptake as a Predictor of Patient Outcome. J Nucl Med 2008;49:1973–9. doi:10.2967/jnumed.108.053397.

[37] Sica GT. Bias in Research Studies. Radiology 2006;238:780–9. doi:10.1148/radiol.2383041109.

[38] Shafiq-ul-Hassan M, Zhang GG, Latifi K, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. Med Phys 2017;44:1050–62. doi:10.1002/mp.12123.

[39] Nyflot MJ, Yang F, Byrd D, et al. Quantitative radiomics: impact of stochastic effects on textural feature analysis implies the need for standards. J Med Imaging (Bellingham, Wash) 2015;2:041002. doi:10.1117/1.JMI.2.4.041002.

Figure Legends

Figure 1: The radiomics workflow.

Abbreviations**:** ICC: intra class coefficient, KM: Kaplan-Meier, McGill: McGill University, n: number, ROC: Receiver Operating Characteristic, TUM: technical University of Munich, UW: University of Washington, Seattle, VOI: volume of interest.

Figure 2: *Radiomics*, *Clinical* and *Combined-Radiomics* prognostic models predicting overall survival (OS), distant (DPFS) and local progression free survival (LPFS).

Receiver operator characteristic curves (ROC) and area under the curve (AUC) values depicting the performance of the *Radiomics* models (black), the *Clinical* models (red) and the *Combined-Radiomics* models (blue) on the training (A,B,C) and validation cohort (D,E,F) for OS (A,D), DPFS (B,E) and LPFS (C,F) dichotomized at two years, respectively.

Figure 3: Patient risk stratification.

Kaplan Meier survival curves for patients' overall survival (OS) (A,D), distant progression free survival (DPFS) (B,E) and local progression free survival (LPFS) (C,F) depicting classifications by the radiomic models on the training (TUM) (A,B,C) and validation cohort (UW) (D,E,F). The median was used as cut-off for dichotomization. Statistical significance of the separation of survival curves was tested using the log-rank test.

Tables

Table 1: Patient demographics, outcome and treatment specifics of included patients.

Institution		TUM	UW	McGill	p-values[1]
Accrual time		2007-2017	2007-2015	n/a	
Total Patients		83 p	87 p	42 p	
Available outcome data	OS/LPFS	74 p (89.2%)	87 p (100%)	42 p (100%)	
	DPFS	76 p (91.6%)	87 p (100%)	42 p (100%)	
Clinical Situation	Primary	74 (89%)	87 p (100%)	42 p (100%)	p=0.002
	Recurrent	9 p (11%)	0 p (0%)	0 p (0%)	
Location	Extremity or trunk	76 p (92%)	86 p (99%)	42 p (100%)	p=0.27
	Abdomen/ retroperitoneal	7 p (8%)	1 p (1%)	0 p (0%)	
Age		m 58 (sd: 16)	m 52.4 (sd:15.8)	m 58.8 (sd:16.6)	p=0.68
Gender	female	36 p (43.4%)	35 p (40.2 %)	24 p (57%)	p=1
	male	47 p (56.6%)	52 p (59.8 %)	18 p (43%)	
T-stadium[2]	1	10 p (12%)	16 p (18.4 %)	n/a	p=0.50
	2	73 p (88%)	70 p (80.5 %)	n/a	
	a	4 p (4.8 %)	5 p (5.8%)	n/a	p=1
	b	79 p (95.2 %)	81 p (93.1%)	n/a	
	unknown	0 p (0%)	1 p (1.1 %)	n/a	
M-stadium[2]	0	70 p (84.3 %)	86 p (98.9%)	n/a	p=1
	1	3 p (3.6 %)	0 p (0%)	n/a	
	X	10 p (12.0 %)	1 p (1.1%)	n/a	
N-stadium[2]	0	73 p (88.0 %)	86 p (98.9%)	n/a	p=0.95
	1	3 p (3.6%)	0 p (0%)	n/a	
	X	7 p (8.4%)	1 p (1.1%)	n/a	
Grading[3]	1	4 p (4.8%)	17 p (19.5%)	4 (10%)	p=1
	2	27 p (32.1%)	28 p (32.2%)	14 (33%)	

	3	49 p (46.4%)	42 p (48.3%)	21 (50%)	
	unclassified	3 p (3.6%)	0 p (0%)	3 p (7%)	
AJCC-Stage	IA	0 p (0%)	5 p (5.7%)	n/a	p<0.0001
	IB	6 p (7.2%)	12 p (13.8%)	n/a	
	IIA	9 p (10.8%)	11 p (12.6%)	n/a	
	IIB	2 p (2.4%)	20 p (23.0%)	n/a	
	III	63 p (75.9%)	38 p (43.7%)	n/a	
	IV	3 p (3.6%)	0 p (0%)	n/a	
	unknown	0 p (0%)	1 p (1.2%)	n/a	
Prognosis					
Median OS		30.1 mo	39.7 mo	24.9 mo	p=1
Median DPFS		17.3 mo	31.5 mo	17.6 mo	p=1
Median LPFS		26.5 mo	31.7 mo	24.0 mo	p=0.49
Therapy information					
Margin-status	positive	16 p (19.3%)	13 p (15.8%)	n/a	p=1
	negative	54 p (65.1%)	69 p (84.1%)	n/a	
	unknown	13 p (15.6%)	5 p (5.7%)	n/a	
Radiotherapy		83p (100%)	87 p (100%)	38 p (90%)	p<0.0001
Total Dose		m 50 Gy (sd: 5.6 Gy)	m 50 Gy (sd: 0.4 Gy)	n/a	p=0.86
Chemotherapy		12 p (14.5%)	49 p (56.3%)	16 p (38%)	p<0.0001

Abbreviations: m: median, McGill: McGill University, mo: months, p: patients, TUM: Technical University of Munich, UW: University of Washington, Seattle, sd: standard deviation.

[1]Categorial variables: Fisher' exact test (2 cohorts) / Chi-square test (3 cohorts) and continuous/Rank variables: Wilcoxon rank sum test (2 cohorts) / Kruskal-Wallis test (3 cohorts), log-rank test for survival, with bonferroni correction for multiple testing

[2]Following AJCC staging system version 7.

[3]According to French Federation of Cancer Centers Sarcoma Group (FFCCS).

Table 2: Predictive performance metrices of radiomic and clinical models.

Model	OS			DPFS			LPFS		
	TUM	UW	McGill	TUM	UW	McGill	TUM	UW	McGill
Radiomic model	**0.72*** (0.61-0.83)	**0.73*** (0.63-0.82)	0.59 (0.39-0.79)	0.64 (0.47-0.70)	**0.68*** (0.58-0.77)	**0.73*** (0.61-0.83)	0.56 (0.40-0.71)	**0.77*** (0.66-0.87)	0.15 (0.04-0.29)
Clinical model	**0.76*** (0.68-0.81)	**0.69*** (0.57-0.80)	n/a	0.46 (0.42-0.53)	**0.66*** (0.55-0.76)	n/a	**0.68*** (0.51-0.81)	0.57 (0.43-0.72)	n/a
Combined model	**0.75*** (0.70 - 0.80)	**0.76*** (0.67-0.84)	n/a	0.60 (0.49-0.68)	**0.68*** (0.55-0.77)	n/a	0.62 (0.47-0.79)	**0.71*** (0.40-0.75)	n/a
AJCC[a] staging	**0.63*** (0.56-0.69)	**0.68*** (0.57-0.77)	n/a	0.55 (0.48-0.61)	**0.66*** (0.54-0.74)	n/a	**0.57*** (0.51-0.63)	0.61 (0.48-0.73)	n/a
Volume	**0.69*** (0.58-0.79)	**0.65*** (0.53-0.76)	0.60 (0.41-0.78)	0.59 (0.49-0.69)	**0.65*** (0.54 -0.75)	**0.67*** (0.54-0.80)	0.61 (0.45-0.75)	0.57 (0.44-0.71)	0.34 (0.19-0.50)
Clinical + volume	**0.78*** (0.74-0.84)	**0.72*** (0.60-0.83)	n/a	0.51 (0.42-0.61)	**0.70*** (0.59-0.80)	n/a	0.63 (0.49-0.75)	0.58 (0.40-0.76)	n/a

Concordance index values for the prediction of overall survival (OS), distant (DPFS) and local progression free survival (LPFS). 95 % confidence intervals are shown in brackets. Predictions significantly different from random are marked with an *.

Abbreviations: McGill: McGill University, TUM: Technical University of Munich, UW: University of Washington

[a] AJCC (American Joint Committee on Cancer) staging manual 7th Edition [1].

Table 3: Cox regression for patients' overall survival and distant progression free survival.

	OS univariate		OS multivariate		DPFS univariate		DPFS multivariate		LPFS univariate		LPFS multivariate	
	HR (95% CI)	p-value [b]	HR (95% CI)	p-value [b]	HR (95% CI)	p-value [b]	HR (95% CI)	p-value [b]	HR (95% CI)	p-value [b]	HR (95% CI)	p-value [b]
TUM - development patient data set												
Radiomics Score	2.5 (1.6-3.7)	0.0002*	3.61 (1.29-10.11)	0.12	2.5 (1.6-3.7)	0.023*	2.66 (1.51-4.68)	0.0058*	0.93 (0.58-1.5)	1	1.4 (0.67-2.9)	1
Age	1.1 (1-1.1)	0.0022*	1.04 (1.00-1.08)	0.464	1 (1-1)	0.66	1 (0.97-1.03)	1	1 (1-1.1)	0.1	1 (0.99-1.1)	1
Grading	2.1 (0.86-5.3)	0.8	0.85 (0.3-2.41)	1	2 (0.95-4.4)	0.55	1.28 (0.54-3.03)	1	3.3 (0.94-11)	0.5	3.5 (0.76-16)	0.85
T-stage	8.5e+07 (0-Inf)	1	3440751.58 (0 - Inf)	1	2 (0.61-6.5)	1	1.59 (0.33-7.69)	1	2.8 (0.37-21)	1	1.9 (0.21-17)	1
T-suffix (a/b)	2.6e+07 (0-Inf)	1	3782697.79 (0-Inf)	1	1.2 (0.16-8.6)	1	0.57 (0.07-4.57)	1	0.62 (0.081-4.7)	1	0.1 (0.0077-1.3)	0.66
N-stage	3.5 (1-12)	0.35	4.12 (0.51-33.4)	01	1.7 (0.4-7.1)	1	1.1e-07 (0-Inf)	1	1.7 (0.22-13)	1	3.2 (0.34-31)	1
M-stage	9.8 (2.1-45)	0.027*	2.73 (0.4-18.69)	1	1.1e-07 (0-Inf)	1	3.9e-08 (0-Inf)	1	1.1e-07 (0-Inf)	1	1.7e-08 (0-Inf)	1
Volume (ml)	1.001 (1.000-1.001)	0.0006*	1	1	1.001 (1.0-1.0001)	0.17	0.99 (0.99-1.00)	1	1 (0.99-1.00)	0.34	1 (0.99-1.0001)	1
UW - validation patient data set												
Radiomics Score	1.9 (1.4-2.6)	0.003*	1.3 (1.1-1.5)	0.0089*	3.3 (1.5-7.7)	0.027*	1.3 (1.1-1.6)	0.0042*	1.4 (1.1-1.7)	0.037*	1.4 (1.1-1.7)	0.025*
Age	1 (1-1.1)	0.58	1 (1-1.1)	0.62	1 (0.99-1)	1	1 (0.99-1)	1	1 (0.98-1)	1	1 (0.98-1.1)	1
Grading	2.7 (1.4-5.5)	0.04*	2.3 (1.2-4.7)	0.11	3 (1.6-5.9)	0.0096*	2.6 (1.3-5.1)	0.04*	1.1 (0.55-2.1)	1	0.92 (0.45-1.8)	1
T-stage	2.3 (0.55-9.9)	1	0.93 (0.21-4.1)	1	1.6 (0.55-4.6)	1	0.54 (0.17-1.7)	1	3.5 (0.45-26)	1	2.4 (0.29-20)	1
T-suffix (a/b)	2.7e+07 (0-Inf)	1	2.1e+07 (0-Inf)	1	7.4e+07 (0-Inf)	1	9.6e+07 (0-Inf)	1	7.4e+07 (0-Inf)	1	3.8e+07 (0-Inf)	1
N-stage	n/a	n/a		n/a	n/a	n/a	n/a		n/a	n/a	n/a	n/a
M-stage	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Volume (ml)	1.001 (1.0000-1.0001)	0.0046*	1.000 (0.9999-1.000	0.69	1.001 (1.0000-1.000	0.0088*	1.0001 (0.9999-	0.33	1.0001 (0.9999-	1	1.0001 (0.9999-	1

	1 (1-1)		1)		1)		1.000 1)		1.000 1)		1.000 1)	
McGill - validation patient set												
Radiomics Score	1.4 (0.79-2.6)	0.48	1.4 (0.64-3.2)	0.76	1.8 (1.1-2.9)	0.026 *	1.3 (0.83-1.9)	0.55	0.11 (0.00043-27)	0.86	0.19 (0.0001-250)	1
Volume	1.000 (0.9999 - 1.0001)	0.86	1.000 (0.9999-1.0001)	1	1.000 (0.9999-1.0001)	0.86	1.000 (0.9999-1.0001)	0.28	1 (0.99-1)	1	1 (0.9999-1.000)	1

Univariate cox regression was performed for the radiomics score and all clinical features. Significant factors are marked by *.

Abbreviations: CI: confidence interval, DPFS: distant progression free survival, HR: hazard ratio, n/a: not available, McGill: McGill University, OS: overall survival, TUM: Technical University of Munich, UW: University of Washington.

[a] In multivariate analysis all patient with non-0 M-status were excluded due to missing variables.

[b] Bonferroni correction was performed for multiple testing.

Bibliography

1. Edge, S. B.; Compton, C. C. The American Joint Committee on Cancer: the 7th Edition of the AJCC Cancer Staging Manual and the Future of TNM. *Ann. Surg. Oncol.* **2010**, *17*, 1471–1474, doi:10.1245/s10434-010-0985-4.

Figures

Figure 1: The radiomics workflow.

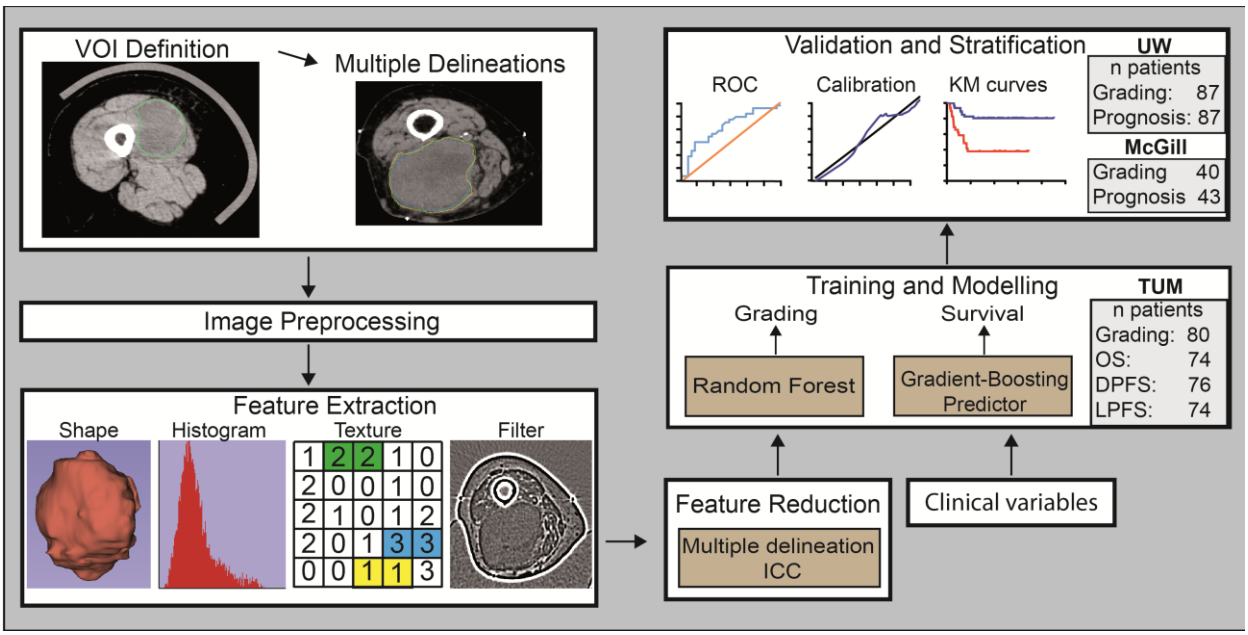


Figure 2. *Radiomics*, *Clinical* and *Combined* prognostic models predicting overall survival (OS), distant (DPFS) and local progression free survival (LPFS).

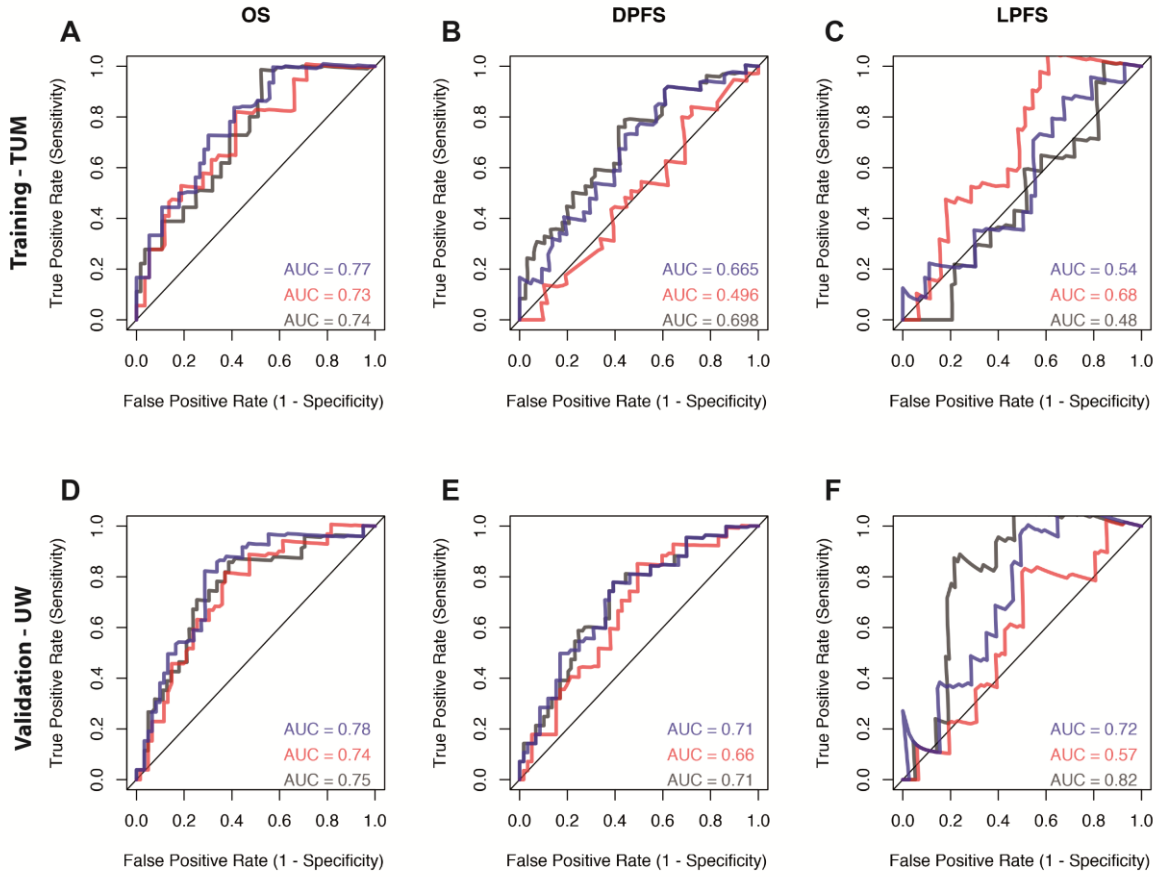
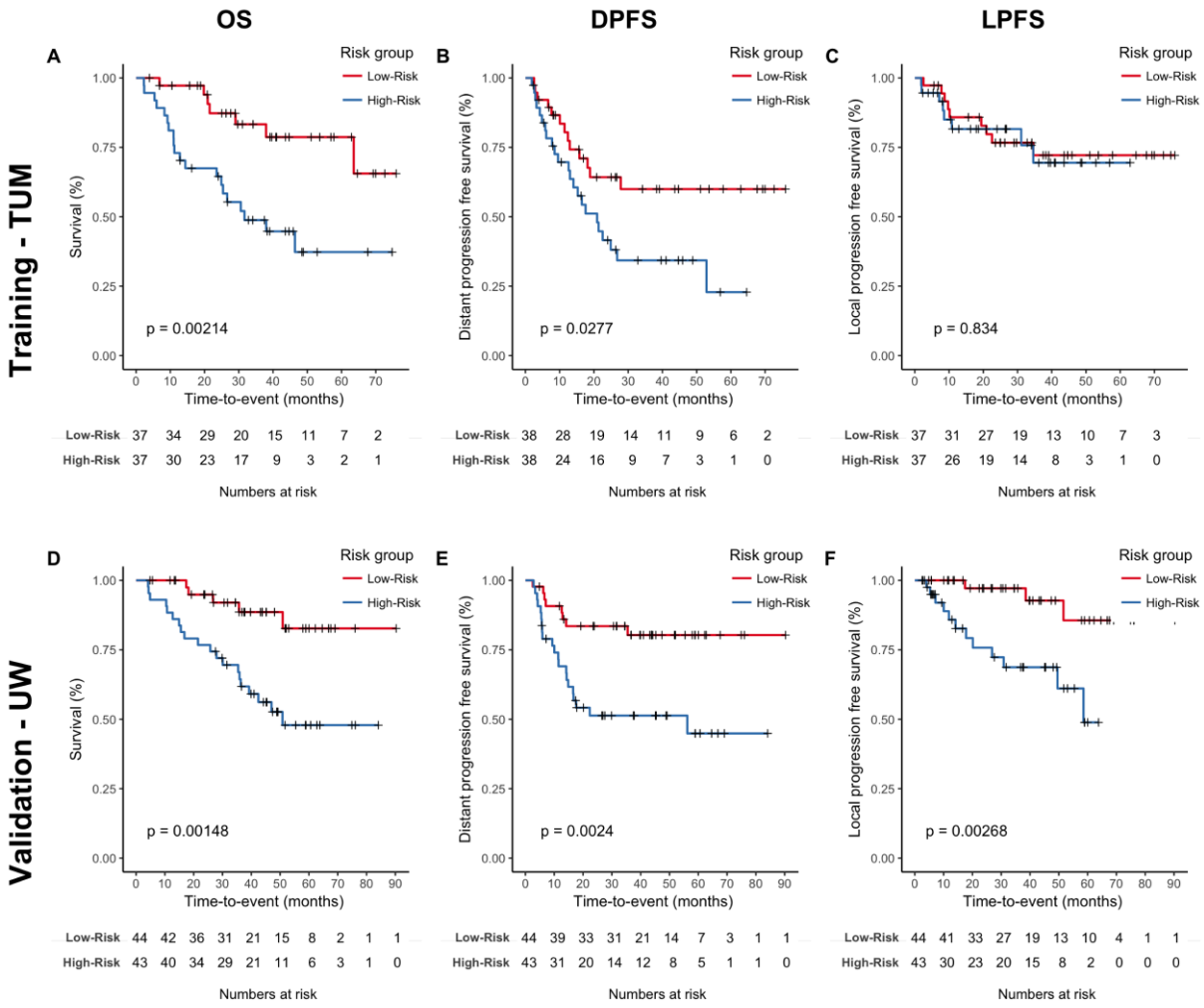


Figure 3: Patient risk stratification.



Conflict of Interest Statement:

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Highlights

1.) CT-based radiomic features have prognostic value in soft tissue sarcoma patients.
2.) CT-based radiomic features differentiate grade 3 STS with moderate performance.
3.) A CT-based radiomic phenotype enables significant patient stratification.
4.) CT-based radiomic phenotypes predict systemic and local progression.