Habitat imaging based on voxel-wise GLCM joint energy clustering: prediction of disease-free survival in localized soft-tissue sarcoma.

Benoît Allignet[1,2], Benjamin Leporq[1], Amine Bouhamama[3], Frank Pilleul[3], Alexandra Meurgey[4], Gualter Vaz[5], Marie-Pierre Sunyach[2], Waisse Waissi[2], Olivier Beuf[1]

[1]CREATIS, CNRS UMR 5220, Inserm U1294, INSA-Lyon, Université Claude Bernard Lyon 1
[2]Department of Radiotherapy, Centre Léon Bérard, Lyon, France
[3]Department of Radiology, Centre Léon Bérard, Lyon, France
[4]Department of Biopathology, Centre Léon Bérard, Lyon, France
[5]Department of Surgical Oncology, Centre Léon Bérard, Lyon, France

Abstract We evaluated if disease-free survival (DFS) of localized soft-tissue sarcoma (STS) patients can be predicted by tumor habitats based on a single voxel-wise radiomics map. The GLCM joint energy map was calculated on fat suppressed T1-weighted contrast-enhanced MRI, performed before neoadjuvant radiotherapy. Habitats were individualized based on k-means clustering method. The median relative volume of habitat 1 (RVH1, reflecting high heterogeneity) was 56.1%, and was used for dichotomization of the sample. Between 2010 and 2020, 149 soft-tissue sarcoma patients were retrospectively included. In this real-life cohort, RVH1 group was independent of MRI acquisition parameters. After a median follow-up of 46.3 months, 4-year DFS was significantly shorter in RVH1-high compared to RVH1-low group (52.5% *vs* 71.5%, p=0.029). In multivariate analysis, RVH1-high and FNCLCC grade 3 remained significantly correlated with shorter DFS (both HR=1.8, p=0.046 and p=0.033, respectively). Habitat imaging based on voxel-wise radiomics may be a valuable tool to evaluate intra-tumor heterogeneity to predict prognosis and/or to personalize treatment.

1 Introduction

Soft-tissue sarcomas (STS) are rare ubiquitous neoplasms with poor prognosis. They present important inter-tumor heterogeneity, from indolent to highly aggressive disease, but also intra-tumor heterogeneity, associating various types of tissues.

Radiomics refers to the process of extracting multiple quantitative imaging biomarkers to use them for outcomes modelling and clinical decision making. This analytical approach usually summarises tumor characteristics into various 1-dimensional (i.e. scalars) features and perform a data-driven approach for outcome prediction. Data are available using various imaging technics applied to various indications, including MRI and STS [1].

New approaches are emerging to consider spatial profiling [2]. Voxel-wise radiomics generate radiomics filtered images and could be more interpretable than 1D radiomics. On the other hand, imaging data can be used to identify tumor sub-regions ("habitats") with different biological properties.

We here propose to combine these habitat imaging and voxel-wise radiomics approach to predict disease-free survival (DFS) in localized STS patients treated by neoadjuvant RT and surgery.

2 Materials and Methods

Patients and magnetic resonance imaging

We retrospectively included all adult patients treated by normofractionated neoadjuvant RT for a localized limb or truncal STS in Centre Léon Bérard (Lyon, France) from January 2010 to November 2020. Exclusion criteria were retroperitoneal STS, bone, Ewing and Ewing-like sarcoma, desmoid tumors, experimental (neo)adjuvant treatment, metastatic disease. Since MRI were performed in various centers, we focused on the only sequence common to all patients: the fat-suppressed T1-weighted gadolinium-enhanced sequence, acquired in 2 or 3 dimensions.

Segmentation and voxel-wise radiomics extraction

Anonymized images were processed on Slicer3D® software. In MRI, magnetic field inhomogeneity and MR coils sensitivity effect can generate a low frequency intensity non-uniformity, called image bias field, that can distort images. Bias field Image bias field was corrected by the retrospective histogram based N4ITK algorithm [3]. Tumors were manually segmented (BA, 5 years of experience). Voxel-wise radiomics were then extracted after z-score normalization, with a fixed bin size of 32 and a kernel radius of 3. B-spline interpolation and nearest neighbor interpolation were used to resample images and segmentation masks, respectively, to 1 mm isotropic voxel size. Radiomics extraction was performed using Python (v3.11.2) and reference packages PyRadiomics (v3.0.1), Numpy (v1.24.2), and SimpleITK (v2.2.1).

K-means clustering-based habitats

The grey-level co-ocurrence matrix joint energy (GLCM-JEnerg) is a measure of homogeneous patterns in the image:

$$GLCM_{JointEnergy} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\left(p(i,j)\right)^2$$

with N_g the discrete number of intensity in the image and p(i,j) the normalized co-ocurrence matrix. We planned to focus on this 3D feature-map since it was previously described as repeatable and reproducible [4], and correlated with local recurrence in glioblastomas [5].

A k-mean clustering method was used on the GLCM-JEnerg map to separate various habitats according to their degree of local heterogeneity. For a number k of cluster (habitat) between 1 and 10, the centroids were calculated to minimize the sum of the squared Euclidean distance

differences between each value and the centroid of the cluster (*within cluster sum of square*, WCSS):

$$WCSS(k) = \sum_{j=1}^{k} \sum_{x_i \in cluster\ j} \|x_i - \bar{x}_j\|^2$$

with $\bar{x}_j$ the centroid of cluster j. The number of clusters for all maps was the optimal number of the majority, determined by the Elbow method.

The habitats were sorted by increasing centroid values (i.e. habitat 1 is the most heterogeneous one), and relative volume occupied by habitat 1 (RVH1) was recorded. The threshold used to dichotomize RVH1 was the overall cohort median value. All this clustering step was performed on MatLab (R2023b).

Clinical outcomes

Primary outcome was disease-free survival (DFS). Secondary outcomes included local recurrence-free survival (LRFS) and predictive factors of RVH1.

Statistical analyses

Categorical variables were reported as counts (percentages) and compared using Chi2 tests. Continuous variables were reported as median (interquartile range) and compared using T-tests or Welch tests (according to the normality of the distributions), or by ANOVA. Median follow-up was determined using the reverse Kaplan Meier method. Survival outcomes were estimated by the Kaplan-Meier method. Prognostic factors were explored in univariate and multivariate Cox proportional hazard models and included RVH1 and usual known factors: age ≥70 years, grade 3 according to fédération nationale des centres de lutte contre le cancer (FNCLCC), deep-seated tumor, size > 5 cm, undifferentiated pleomorphic sarcoma or myxofibrosarcoma histotypes. Two-sided p-value <0.05 was considered significant. Statistical analyses were performed using R software version 4.2.3 (R Foundation for Statistical Computing, Vienna, Austria).

3 Results

We retrospectively included 149 STS patients. Their clinical characteristics and the MRI acquisition parameters are summarized in the Table 1.

Table 1. Patients' characteristics

Characteristics	Overall (n=149)
Age	61 (48-73)
Sexe	
Male	72 (48.3)
Female	77 (51.7)
Deep-seated lesion	138 (92.6)
Location	
Extremities	118 (79.2)
Trunk	31 (20.8)
Histotype	
UPS	52 (34.9)
Myxoid liposarcoma	23 (15.4)
Myxofibrosarcoma	22 (14.8)
Dediff./pleo. liposarcoma	17 (11.4)
Synovial sarcoma	14 (9.4)
Others	21 (14.1)
FNCLCC grade	
Grade 1-2	89 (62.7)
Grade 3	50 (35.2)
Neoadjuvant chemotherapy	40 (26.8)
Number of cycles of NAC	4 (3-6)
Radiotherapy technique	
Tomotherapy	67 (45)
VMAT	39 (26.2)
3D-CRT	43 (28.9)
Total dose	50 (50-50)
Overall treatment time	36 (35-39)
Volumes of radiotherapy (cc)	
GTV	263 (115-500)
CTV	957 (457-1387)
PTV	1183 (599-1741)
Surgical margins	
R0	123 (82.6)
R1	26 (17.4)
Magnetic field of the MRI	
1.5 T	118 (79.2)
3.0 T	31 (20.8)
Dimension of acquisition	
2D	81 (54.4)
3D	68 (44.6)
Type of fat-suppressed T1-CE sequence	
Gradient echo sequence	63 (42.3)
Flip angle	11 (10-11)
Echo time	2.46 (2.39-2.46)
Repetition time	6.38 (5.55-6.79)
Spin echo sequence	86 (57.7)
Echo time	11 (9.5-13)
Repetition time	586.2 (531.25-719.5)
Method of fat suppression	
Fat-water decomposition	83 (55.7)
Fat saturation	60 (40.3)
Inversion-recovery imaging	6 (4)

Note:
Continuous variables: median (IQR)
Categorial variables: n (%)

Characteristics	Overall (n=149)

Abbreviations: UPS, undifferentiated pleomorphic sarcoma; dedif./pleo., dedifferentiated/pleomorphic; FNCLCC, fédération nationale des centres de lutte contre le cancer; NAC, neoadjuvant chemotherapy; VMAT, volumetric modulated arctherapy; T1-CE, T1-weighted contrast-enhanced sequence

The optimal number k of habitats was 4 for most tumors (n=111; k=3 for n=37 and k=5 for n=1), with a median centroid of habitat 1 of 0.021 (IQR 0.016-0.029).
The median RVH1 was 56.1% (49-65) (Figure 1). The RVH1 group was not related to the MRI magnetic field strength, type of sequence, dimension of acquisitions (all p>0.96) nor to the method of fat suppression (p=0.23). It was not either related to histological subtype (p=0.09), nor FNCLCC grade 3 (p=0.2), tumor absolute volume (p=0.17) or predictive of the percentage of viable cells on surgical specimen (p=0.9).

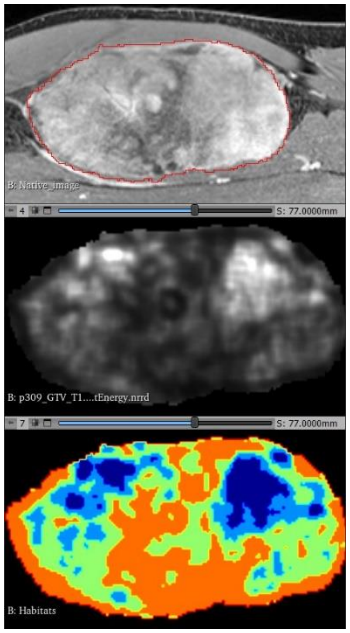

Figure 1. Native T1-weighted gadolinium enhanced Dixon-water image (up) with its GLCM -Joint Energy corresponding map (middle) and the habitat maps (down). The gross tumor volume is segmented in red. The habitat 1 to 4 are depicted from orange to dark blue, respectively.

After a median follow-up of 46.3 months (95%CI 33.1-70.1), 4-year DFS was 71.5% and 52.5% in RVH1-low and RVH1-high group, respectively (Figure 2), with 4-year LRFS of 77.3% and 65%, respectively (Figure 3). In univariate analysis, only grade 3 was significantly correlated with LRFS (HR 2.3, 95%CI 1.2-4.2, p=0.01). In both uni- and multivariate analyses, DFS was significantly related to FNCLCC grade 3 and RVH1 (HR 1.8, 95%CI 1.1-3.1, p=0.033; HR 1.8, 95%CI 1.01-3.1, p=0.046, respectively).

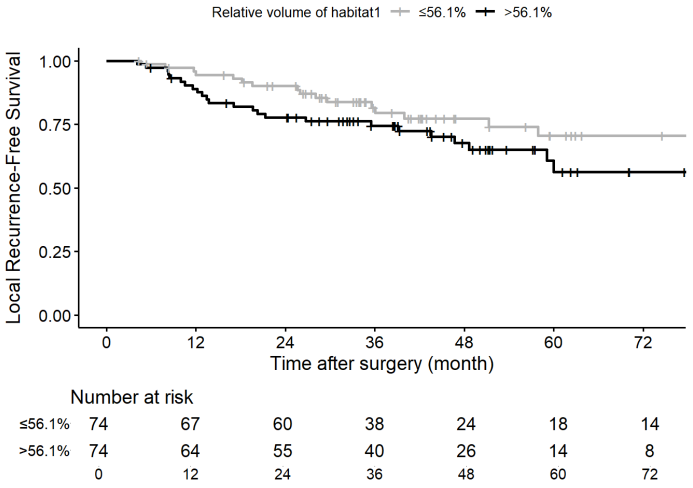

Figure 2. Local recurrence-free survival according to relative volume of habitat 1

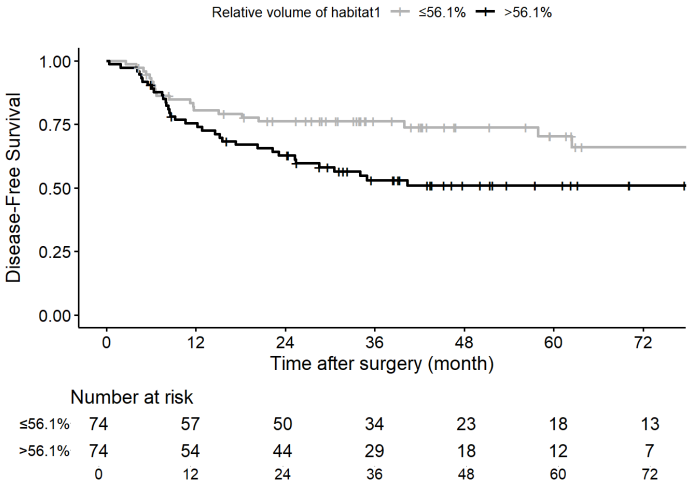

Figure 3. Disease-free survival according to relative volume of habitat 1

As a sensitivity analysis, we divided the cohort into 3 groups according to the terciles of the RVH1 distribution. The DFS tended to shorten with the increase of RVH1, with 4 year-DFS of 75% *vs* 60.9% *vs* 50.8% (Figure 4).

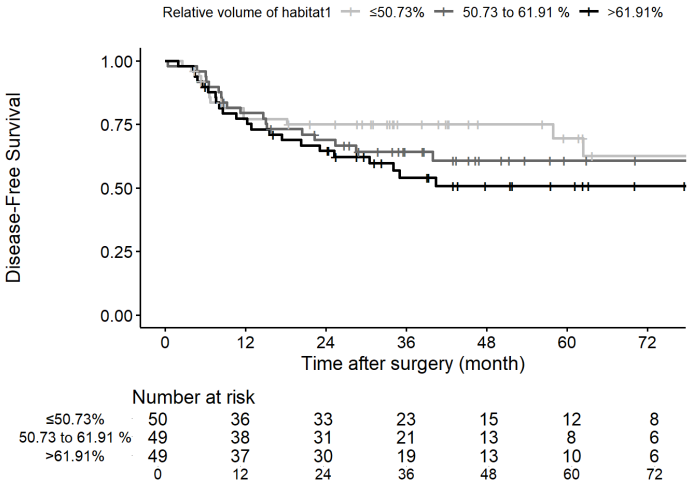

Figure 4. Disease-free survival according to relative volume of habitat 1, divided into 3 groups.

4 Discussion

We here report an original methodology of habitats imaging based on fat suppressed T1-weighted contrast-enhanced sequence voxel-wise radiomics. GLCM-JEnerg RVH1-low group present a significantly longer DFS, with a clinically meaningful difference.

This highlights the potential interest of voxel-wise radiomics either to identify tumor subclones or to predict patient prognosis.

Nevertheless, some refinements of the method may be performed. First, the repeatability of the habitats may be verified, for example by adding Rician noise in native images and process them back. As a retrospective analysis, the reproducibility of the method when performing other MR acquisitions cannot be confirmed. In contrast, an external cohort may be used to reproduce and validate the whole processes and results.

MR acquisition parameters were variable between patients and could thus introduce bias, particularly using vowel-wise radiomics [4]. Nevertheless, various steps limited this: intensity non-uniformity was corrected, native images were standardized, and a clustering was individually performed. Furthermore, the whole process was not data- but hypothesis-driven, with interpretable intermediate results, and is applied on real-world data acquired on an important cohort of patients, with sufficient follow-up, presenting rare disease managed in an expert center.

5 Conclusion

In this real-world cohort of localized STS, habitat imaging based on a single voxel-wise radiomic map was independently correlated with DFS. This kind of approach could be a valuable tool to integrate intra-tumor heterogeneity into RT personalization.

Acknowledgements

With financial support from ITMO Cancer of Aviesan within the framework of the 2021-2030 Cancer Control Strategy, on funds administered by Inserm, and from the committee 69 of the Ligue Nationale Contre le Cancer. This work was also performed within the framework of the the LabEx PRIMES (ANR-11-LABX-0063) as well as SIRIC LyriCAN+ grant (INCa INSERM DGOS 12563).

References

[1] Spraker MB, Wootton LS, Hippe DS, Ball KC, Peeken JC, Macomber MW, et al. MRI Radiomic Features Are Independently Associated With Overall Survival in Soft Tissue Sarcoma. Adv Radiat Oncol 2019;4:413–21. DOI: 10.1016/j.adro.2019.02.003

[2] Napel S, Mu W, Jardim-Perassi B V., Aerts HJWL, Gillies RJ. Quantitative imaging of cancer in the postgenomic era: Radio(geno)mics, deep learning, and habitats. Cancer 2018;124:4633–49. DOI: 10.1002/cncr.31630

[3] Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: Improved N3 bias correction. IEEE Trans Med Imaging 2010;29:1310–20. DOI: 10.1109/TMI.2010.2046908.

[4] Bernatowicz K, Grussu F, Ligero M, Garcia A, Delgado E, Perez-Lopez R. Robust imaging habitat computation using voxel-wise radiomics features. Sci Rep 2021;11. DOI: 10.1038/s41598-021-99701-2

[5] Chougule T, Gupta RK, Saini J, Agrawal S, Gupta M, Vakharia N, et al. Radiomics signature for temporal evolution and recurrence patterns of glioblastoma using multimodal magnetic resonance imaging. NMR Biomed 2022;35. DOI: 10.1002/nbm.4647