The discriminative power and stability of radiomics features with CT variations: Task-based analysis in an anthropomorphic 3D-printed CT phantom

Abstract

Objectives: To determine the stability of radiomics features against Computer Tomography parameter variations and to study their discriminative power concerning tissue classification using a 3D-printed CT phantom based on real patient data.

Materials and Methods: A radiopaque 3D phantom was developed using real patient data and a potassium iodide solution paper printing technique. Normal liver tissue and 3 lesion types (benign cyst, hemangioma and metastasis) were manually annotated in the phantom. The stability and discriminative power of 86 radiomics features was assessed in measurements taken from 240 CT series with 8 parameter variations of reconstruction algorithms, reconstruction kernels, slice thickness and slice spacing. Pairwise parameter group and pairwise tissue class comparisons were performed using Wilcoxon signed-rank tests.

Results: In total, 19,264 feature stability tests and 8,256 discriminative power tests were performed. The 8 CT parameter variation pairwise group comparisons had statistically significant differences on average in 78/86 radiomics features. On the other hand, 84% of the univariate

radiomics feature tests had a successful and statistically significant differentiation of the 4 classes of liver tissue. The 86 radiomics features were ranked according to the cumulative sum of successful stability and discriminative power tests.

Conclusions: The differences in radiomics feature values obtained from different types of liver tissue are generally greater than the intra-class differences resulting from CT parameter variations.

Keywords: radiomics, texture analysis, anthropomorphic phantom, computed tomography, quantitative biomarkers

1 Introduction

Radiomics can perform a comprehensive and automated quantification of phenotypic disease characteristics on medical imaging [1, 2, 3]. When radiomics features are combined with other data sources, such as structured clinical data, or genomics data, they can produce robust evidence for clinical-decision support systems [4]. Particularly in oncological studies and tumour characterisation, they have outperformed visual assessment and provide high potential for personalised medicine [5].

One major challenge that still needs to be overcome for the clinical implementation of this technology, is the variation of radiomics feature values caused by changes in the data acquisition and image reconstruction [6, 7, 8]. Several studies have shown that the stability of radiomics features in computed tomography images can be severely impacted by intra- and inter-scanner differences [9, 10, 11]. To minimize the risk of using unstable radiomics features treatment site-specific and time-, scanner-, and imaging protocol-controlled test-retest analyses have been proposed [12]. In recent years, several small studies have carried out test-retest studies with

multiple patients to assess this challenge [13, 11, 14, 15]. However, Computed Tomography (CT) test-retest studies evaluating radiomics feature stability entail ethical concerns of recruiting a large cohort of patients with multiple exposures to radiation and are thus limited to small sample sizes. To address the latter, recent studies used phantoms allowing extensive radiation exposure in highly controlled acquisition conditions [9, 16, 17, 18, 19, 20].

Phantoms aim to mimic human and tumour tissue contrast and attenuation to assess the performance of medical devices in imaging, nuclear medicine and radiation therapy [21]. They are mainly used by hospital imaging departments for regular device calibration. One disadvantage of most commercially available phantoms is their geometric simplicity and uniform backgrounds due to the use of homogeneous materials [9, 22] with very limited texture. Recently, an innovative aqueous potassium iodide solution-based paper printing technology to create realistic 3D CT phantoms based on real CT-acquired patient data was developed [23]. These anthropomorphic radiopaque phantoms include high precision anatomic details and attenuation properties. When compared to patient-based test-retest approaches, phantoms also simplify image registration of different acquisitions. Anthropomorphic phantoms are therefore ideally suited to study the impact of CT acquisition and reconstruction parameters based on realistic densities and texture patterns of human tissue.

Another limitation of most previous studies investigating the effect of CT parameters on radiomics measurements is a focus on parameter stability alone [16, 20], neglecting the discriminative power of the features given a controlled detection or classification task. As long as intra-class variations resulting from CT parameters remain significantly smaller than inter-class variations of different lesions, the consequences of the former are limited. Homayounieh *et al.* had previously evaluated the accuracy of radiomics in differentiating diffuse liver diseases on non-contrast CT [24]. Rotzinger *et al.* investigated the variability in diagnostic information from 68 different CT units in the detection of focal liver lesions using an anthropomorphic abdominal phantom [25]. In the context of radiomics features, Caramella *et al.* also focused on stability and discriminative power

[19]. However, they used a nonanthropomorphic synthetic phantom where the discrimination tasks were not realistic regarding actual use of radiomics biomarkers (*e.g.* normal versus tumoural tissue).

In this paper we present a study of the stability and discriminative power of radiomics features using a 3D printed texture phantom of the thorax and abdomen. A univariate analysis designed to better understand the influence of CT variations on individual radiomics features is performed toward the translation of these advanced image computing techniques to clinical practice. When compared to previous work, our approach is novel in two aspects: (i) high sample size and realistic image acquisitions using an anthropomorphic radiopaque phantom and (ii) the consideration of combined stability and discriminative power of radiomics features.

2 Materials and Methods

2.1 Phantom

A realistic radiopaque 3D phantom designed to simulate clinical imaging in oncology was built showing real patient texture data of the thorax and the abdomen, including lesions and tumours. The customised phantom was manufactured by printing real patient CT data with an aqueous potassium iodide solution on paper and subsequently joining the sheets together [23]. Tissue equivalent attenuation at a defined energy spectrum was calibrated at 120 kVp (PhantomX, www.thephantomx.com). The phantom comprises two separate sections: a half-mirrored lung tumour section, which was not used in this study, and an abdominal section with one pathology proven liver metastasis from a colon carcinoma (see details in Supplementary Figure S1 [online]). In Fig. 1) the phantom is shown together with manually annotated Regions Of Interest (ROI) in the liver. The lung tumour section is derived from a non-small-cell lung carcinoma patient (PAT1) that was made publicly available to serve as radiomics phantoms [4], including for the Image Biomarker Standardisation Initiative (IBSI) [26].

2.2 CT acquisitions

The phantom was imaged with a Siemens SOMATOM Definition Edge (Siemens Healthineers, Erlangen, Germany) CT scanner. To define the acquisition and image reconstruction parameters, a survey of clinical CT protocols was done including 9 radiological institutes. Portal venous contrast phase thorax/abdomen protocols used for clinical indications of tumour or infectious foci were considered in the survey. All the CT scans in this study were acquired with the same acquisition parameters. The tube voltage was 120 kVp, the helical pitch factor was 1.0, the rotation time was 0.5 s and the tube current time product was 147 mAs (no automatic tube current modulation), resulting in a volume computed tomography dose index (CTDIvol) of approximately 10 mGy.

The following image reconstruction parameters were varied in this study: reconstruction algorithm (Iterative Reconstruction, IR or Filtered Back Projection, FBP), reconstruction kernel (two standard soft tissue kernels per algorithm), slice thickness in mm (1, 1.5, 2, 3) and slice spacing in mm (0.75, 1 and 2). Series reconstructed with an IR algorithm used an ADvanced Modeled Iterative REconstruction (ADMIRE) at strength level 3. In total, eight groups of parameter variations were selected for this study to assess their impact on classic radiomics features (see Table 1). These reconstruction parameter settings are typical for clinical protocols in thoracic and abdominal oncology.

Two approaches were used during the CT scanning of the phantom. First, twenty repetition scans (=acquisitions) were performed without re-positioning of the phantom. Then, ten repetitions were performed with re-positioning of the phantom between each measurement. For each acquisition, the 8 reconstructions listed in Table 1 were performed. In total, 30 CT series (from 30 distinct acquisitions) were available for each of the 8 parameter variations selected for this study. All series were then exported in the Digital Imaging and Communications in Medicine (DICOM) format.

2.3 Regions of interest and registration

Six 3D ROIs were manually annotated by a board-certified radiologist using a phantom series with 2 mm slice thickness and 1 mm spacing. Normal liver tissue and three different types of liver lesions were included during the annotation process resulting in two normal liver tissue regions, two cysts, a hemangioma and a liver metastasis from a colon carcinoma.

All CT volumes were registered to a common space using an automated rigid registration approach from the Elastix toolbox [27]. The phantom in the volumes was treated as a rigid body, allowing translation and rotation, thus matching the manually annotated ROIs to the same phantom location in all available CT series.

2.4 Radiomics feature extraction

A total of 86 radiomics features were extracted in 3D from the manually segmented ROIs using the open source Pyradiomics (version 3.0) python toolkit [2]. The radiomics feature categories include First-Order statistics (firstorder, n=18), Grey Level Co-Occurrence Matrices (GLCM, n=22), Grey Level Dependence Matrices (GLDM, n=14), Grey Level Run Length Matrices (GLRLM, n=16), and Grey Level Size Zone Matrices (GLSZM, n=16). No shape features were considered in this study since the features were extracted from the same ROIs in the registered CT volumes and are therefore not varying across acquisitions. Definitions for the radiomics features are available in the Pyradiomics documentation online (PyRadiomics, Radiomics features, pyradiomics.readthedocs.io/en/latest/features.html).

2.5 Statistical analysis

The Wilcoxon signed-rank test is a non-parametric test that does not require the assumption of normality. This test was selected to assess the stability and discriminative power of isolated radiomics features in two separate univariate analyses. A two-tailed test with a p-value < 0.05 was considered statistically significant. All statistical analyses were performed using the Scikitlearn and Scipy packages implemented for Python software (version 3.7) [28].

2.5.1 Stability of radiomics features (intra-class variation)

For the stability tests, a tissue class is fixed and each radiomics feature set includes the measurements obtained from the 30 CT series sharing the same parameter settings. Pairwise comparisons between the eight parameter variation groups are then performed with Wilcoxon signed-rank tests. This process is repeated for all available tissue classes while all other CT parameters are kept constant. A univariate feature stability test is considered positive if the difference between the parameter variation groups is not statistically significant.

2.5.2 Discriminative power of radiomics features (inter-class variation)

In the discriminative power tests, a CT parameter setup is kept constant while the feature values obtained from a specific ROI class (*i.e.* normal liver tissue, cyst, hemangioma and liver metastasis) are compared in pairwise tests to the values from the remaining classes. These comparisons are performed for each of the 8 parameter variation groups. A discriminative power test for radiomics features is considered positive if a pairwise comparison has statistical significance.

3 Results

In a preliminary analysis, no statistical significance was found between the CT series without phantom repositioning and those acquired with re-positioning of the phantom between each

acquisition. Therefore, the 30 CT series sharing the same CT parameter settings were then considered as a single group of feature value measurements.

3.1 Feature stability

In total, 19,264 feature stability tests were performed: 8 parameter variations × 7 pairwise comparisons × 4 liver tissue classes × 86 radiomics features. The confusion matrix resulting from the feature stability tests is shown in Figure 2. The matrix shows the mean proportion of radiomics features without statistically significant differences for pairwise comparisons between the 8 parameter variation groups defined in Table 1. Only in 15% of these feature stability tests, the CT parameter variations had no significant impact on the radiomics features values. Moreover, 78 out of the 86 radiomics features had, on average, statistically significant differences in the 8 CT parameter variation pairwise group comparisons. It can be noted that the parameter groups with the highest proportion of significant differences to the other groups have a slice thickness of 3 mm. In Figure 3, a qualitative evaluation of some of the liver texture patterns (*i.e.* cyst and liver metastasis) is shown for 4 groups of CT parameter variations (*i.e.* Groups 1,3,6 and 8).

3.2 Discriminative power

Regarding the task-based discriminative power of radiomics features, 8,256 tests were performed: 4 tissue classes × 3 pairwise comparisons to the other classes × 8 parameter variations × 86 radiomics features. All of the independent radiomics features had at least 40% of successful discriminative power tests. In particular, 84% of the univariate discriminative power tests resulted in a statistically significant differentiation of the 4 classes of liver tissue. In Figure 4, a plot of the 86 radiomics features is shown with the percentage of positive stability tests on the *x*-axis and the percentage of positive discriminative power tests on the *y*-axis.

3.3 Data projection and feature ranking

Figure 5 shows a principal component analysis visualisation of the 86 radiomics features from all 240 available CT series. After visually inspecting the data projected in the feature space of the first two principal components, the differences between the four ROI classes (inter-class variation) are larger than all CT parameter variations (intra-class variation). Moreover, ROIs from the normal liver tissue class (green), are closer in the feature space than those from the other classes. The two measurements inside the benign cyst ROIs resulted in a larger variation, with cyst_2 measurements appearing also close to the measurements obtained from the metastasis ROI. The visual texture patterns and Hounsfield Unit intensity values within the cysts are more diverse than those in the other ROIs, thus generating these larger distributions in the feature space (for more details see Supplementary Figure S2 [online]). Nevertheless, all four classes remain linearly separable despite the CT parameter variations.

The radiomics features were then ranked based on the sum of the percentages of successful stability and discriminative power tests. The top ten radiomics features with best performance and their corresponding percentages of successful tests are shown in Table 2 (full table showing the results for the 86 radiomics features available [online], Supplementary Table S1).

4 Discussion

To the best of our knowledge, this is the first task-based radiomics study targeting the stability and discriminative power of radiomics features using a realistic 3D phantom. The univariate differences with statistical significance are measured for 86 common radiomics features with four types of CT parameter variations (reconstruction algorithm, reconstruction kernel, slice thickness and slice spacing) in a controlled setup. Normal liver tissue regions and 3 types of liver lesions are considered. In this study, both the stability of radiomics features, together with their

discriminative power, are presented in a task-based performance evaluation using real, not simulated, 3D liver tissue patterns.

According to the results from the feature stability tests, changes in CT parameter settings with thin slices, *e.g.* 1-2mm, have fewer features impacted than changes from 1-2mm to thicker slices *i.e.* 3mm. In addition, CT images with thin slices contain more detailed texture information and have been associated with better diagnostic performance for some tasks [7]. Also, when stability studies use real patient data, slice thickness seems to be a more influencing factor than the reconstruction algorithm and reconstruction kernel [11]. Density phantom studies with simple backgrounds do not share this characteristic [9].

The results from this study show that even when the majority of radiomics features had low stability for CT parameter variations, as has been previously shown in other studies [8, 10, 11, 12], the discriminative power remains markedly high in the task of differentiating between 4 tissue classes. While the fraction of significant stability and discriminative power tests depends on the number of repetition acquisitions, the combination of stability and discriminative power results can be used to rank the radiomics features. A detailed analysis of the radiomics features with the best performance both in stability and discriminative power could be used to define guidelines for feature inclusion in future CT imaging studies. Although some compensation methods have been proposed to counteract the effect of the parameter and inter-scanner variations in radiomics studies [29], the feature values were not normalised for these tests. Any normalisation step is highly dependent on the samples and range of values considered for the normalisation. Moreover, the interpretability of the radiomics features is facilitated when their original computation values are respected. In future work we will expand our study with multi-site, different vendor comparisons, as well as the impact of other CT scanning parameters (*e.g.* dose, tube voltage). Finally, the stability of other quantitative biomarkers, *i.e.* deep features, could be further studied to limit the variation due to scanners and CT protocols [31].

Even though the realistic phantom offers many advantages with respect to other phantoms available, a range of low Hounsfield Units (HU) cannot be achieved by the production technique of this phantom, with a limit around -100 HU. However, this did not impact our study, as in the range of HU > 0 (*e.g.* liver tissue) the structures are presented with a high resolution and realistic contrast. Overcoming problems resulting from the 3D printing of realistic relative HU values was also a challenge in the phantom study from Holmes et al. [30]. Given the flexibility and increasing availability of 3D printing, it can be expected that further development of these techniques will be investigated in future studies [21].

In conclusion, CT data can have multiple scanner variations and parameters during the acquisition and reconstruction process that can affect the values of radiomics features. Understanding the scale of changes produced from variations in the acquisition and reconstruction process can help to better interpret the analysis of radiomics studies and in case of retrospective studies the inclusion or exclusion of patients. Identifying features that are not only stable between different scanner settings but that have also a high discriminative power for a specific clinical task should be a leading factor when developing personalised medicine approaches with these biomarkers, thus limiting unwanted influences on the results.

5 References

[1] Gillies RJ, Kinahan PE, and Hricak, H. Radiomics: images are more than pictures, they are data. Radiology. 2016; 278(2):563–577.

[2] van Griethuysen JJ, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. Cancer Research. 2017; 77: e104– e107.

[3] Ibrahim A, Primakov S, Beuque M, et al. Radiomics for precision medicine: current challenges, future prospects, and the proposal of a new framework. Methods. 2020; in press.

[4] Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nature reviews Clinical oncology. 2017; 14(12): 749–762.

[5] Yip SS, Aerts HJ. Applications and limitations of radiomics. Physics in Medicine & Biology. 2016; 61: R150.

[6] Mackin D, Fave X, Zhang L, et al. Measuring CT scanner variability of radiomics features. Investigative Radiology. 2015; 50:757.

[7] He L, Huang Y, Ma Z, et al. Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule. Scientific reports. 2016; 6:34921.

[8] Berenguer R, Pastor-Juan MdR, Canales-Vázquez J, et al. Radiomics of CT features may be nonreproducible and redundant: Influence of CT acquisition parameters. Radiology. 2018; 288:407–415.

[9] Solomon J, Samei E. Quantum noise properties of CT images with anatomical textured backgrounds across reconstruction algorithms: FBP and SAFIRE. Medical Physics. 2014; 41:091908.

[10] Traverso A, Wee L, Dekker A, et al. Repeatability and reproducibility of radiomic features: a systematic review. International Journal of Radiation Oncology, Biology, Physics. 2018; 102:1143–1158.

[11] Meyer M, Ronald J, Vernuccio F, et al. Reproducibility of CT radiomic features within the same patient: influence of radiation dose and CT reconstruction settings. Radiology. 2019; 293:583–591.

[12] van Timmeren JE, Leijenaar RT, van Elmpt W, et al. Test–retest data for radiomics feature stability analysis: generalizable or study-specific? Tomography. 2016; 2:361.

[13] Solomon J, Mileto A, Nelson RC, et al. Quantitative features of liver lesions, lung nodules, and renal stones at multi–detector row CT examinations: dependency on radiation dose and reconstruction algorithm. Radiology. 2016; 279(1): 185–194.

[14] Prayer F, Hofmanninger J, Weber M, et al. Variability of computed tomography radiomics features of fibrosing interstitial lung disease: A test-retest study. Methods. 2020; S1046-2023(20): 30194-8.

[15] Hoebel KV, Patel JB, Beers A, et al. Radiomics Repeatability Pitfalls in a Scan-Rescan MRI Study of Glioblastoma. Radiology: Artificial Intelligence. 2020; e190199.

[16] Shafiq-ul-Hassan M, Zhang GG, Latifi K, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. Medical Physics. 2017; 44:1050–1062.

[17] Larue RTHM, van Timmeren JE, de Jong EEC, et al. Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses: a comprehensive phantom study. Acta Oncologica. 2017; 56(11):1544-1553.

[18] Ger RB, Zhou S, Chi P-CM, et al. Comprehensive investigation on controlling for CT imaging variabilities in radiomics studies. Scientific reports. 2018; 8(1):1-14.

[19] Caramella C, Allorant A, Orlhac F, et al. Can we trust the calculation of texture indices of CT images? A phantom study. Medical Physics. 2018; 45: 1529–1536.

[20] Mackin D, Ger R, Dodge C, et al. Effect of tube current on computed tomography radiomic features. Scientific reports. 2018; 8:1–10.

[21] Tino R, Yeo A, Leary M, et al. A systematic review on 3D-printed imaging and dosimetry phantoms in radiation therapy. Technology in cancer research & treatment. 2019; 18:1533033819870208.

[22] Craft DF, Howell RM. Preparation and fabrication of a full-scale, sagittal-sliced, 3D-printed, patient-specific radiotherapy phantom. Journal of applied clinical medical physics. 2017; 18:285–292.

[23] Jahnke P, Limberg FRP, Gerbl A, et al. Radiopaque three-dimensional printing: a method to create realistic CT phantoms. Radiology. 2017; 282:569–575.

[24] Homayounieh F, Saini S, Mostafavi L, et al. Accuracy of radiomics for differentiating diffuse liver diseases on non-contrast CT. International Journal of Computer Assisted Radiology and Surgery. 2020; 15(10):1727–1736.

[25] Rotzinger DC, Racine D, Beigelman-Aubry C, et al. Task-based model observer assessment of a partial model-based iterative reconstruction algorithm in thoracic oncologic multidetector CT. Scientific reports. 2018; 8:1–12.

[26] Zwanenburg A, Vallières M, Abdalah MA, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. Radiology. 2020; 295:328–338.

[27] Klein S, Staring M, Murphy K, et al. Elastix: a toolbox for intensity-based medical image registration. IEEE transactions on medical imaging. 2009; 29:196–205.

[28] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. Journal of machine learning research. 2011; 12, 2825–2830.

[29] Orlhac F, Frouin F, Nioche C, et al. Validation of A Method to Compensate Multicenter Effects Affecting CT Radiomics. Radiology. 2019; 291: 53–59.

[30] Holmes RB, Negus IS, Wiltshire SJ, et al. Creation of an anthropomorphic CT head phantom for verification of image segmentation. Medical Physics. 2020; 47:2380-2391.

[31] Andrearczyk V, Depeursinge A, Müller H. Neural Network Training for Cross-Protocol Radiomic Feature Standardization in Computed Tomography. Journal of Medical Imaging. 2019; 6:024008.

Tables

Table 1. Summary of CT reconstruction parameter variations in this study. FBP: Filtered Back Projection, IR: Iterative Reconstruction, ASA: Advanced Smoothing Algorithm.

Group	Reconstruction algorithm	Reconstruction kernel	Slice thickness (mm)	Slice spacing (mm)
1	FBP	B26f medium smooth ASA	1	0.75
2	FBP	B30f medium smooth	1.5	1
3	FBP	B30f medium smooth	2	1
4	FBP	B30f medium smooth	3	2
5	IR	I26f medium smooth ASA	1	0.75
6	IR	I30f medium smooth	1.5	1
7	IR	I30f medium smooth	2	1
8	IR	I30f medium smooth	3	2

Table 2. Top ten radiomics features when considering both the univariate stability and discriminative power for differentiating between 4 types of liver tissue (normal liver tissue, benign cysts, hemangioma and liver metastasis) with 8 CT parameter variations. firstorder: First-Order statistics, glcm: Grey Level Co-occurrence Matrix, gldm: Grey Level Dependence Matrix, glszm: Grey Level Size Zone Matrix, glrlm: Grey Level Run Length Matrix.

Features	Stability (%)	Discriminative power (%)	Sum
original_firstorder_Median	65.2	100.0	165.2
original_firstorder_Mean	53.6	100.0	153.6
original_gldm_LargeDependenceHighGreyLevelEmphasis	67.9	81.3	149.1
original_glcm_ClusterShade	56.3	89.6	145.8
original_gldm_SmallDependenceLowGreyLevelEmphasis	54.5	85.4	139.9
original_glcm_InverseDifferenceMomentNormalised	52.7	83.3	136.0
original_firstorder_Skewness	51.8	83.3	135.1
original_glszm_SmallAreaLowGreyLevelEmphasis	67.0	66.7	133.6
original_glrlm_LongRunHighGreyLevelEmphasis	36.6	93.8	130.4
original_glszm_LargeAreaHighGreyLevelEmphasis	25.0	100.0	125.0

Figure legends

Figure 1. Anthropomorphic radiopaque 3D phantom and manually annotated ROIs. A. Photo of the realistic 3D printed phantom set in a scanner. B. Coronal view from a CT volume of the phantom showing the mirrored thorax section (top) and an abdominal section (bottom). C and D. Two axial sample views displaying the six manually annotated ROIs in the liver: normal liver tissue (green), a liver metastasis from a colon carcinoma (red), two benign cysts (blue) and a hemangioma (yellow).

Figure 2. Univariate stability of radiomics features using the proposed pairwise comparisons between the 8 groups of parameter variations. The average percentage of radiomics features that showed a significant difference ($p < 0.05$) in the measurements of 6 liver ROIs obtained from sets of 30 CT series reconstructed with each of the corresponding parameter variations.

Figure 3. Liver texture patterns (cyst and metastasis) in four different CT parameter variations: filtered back projection with slice thickness of 1 mm (Group 1), filtered back projection with slice thickness of 2 mm (Group 3), iterative reconstruction with slice thickness 1.5 mm (Group 6) and iterative reconstruction with slice thickness 3 mm (Group 8).

F**igure 4.** Scatter plot representation of all 86 radiomics features in terms of the percentage of successful stability (*x*-axis) and discriminative power (*y*-axis) tests between pairwise comparisons of the 8 parameter variation groups.

Figure 5. Principal component visualisation for 1440 samples corresponding to 30 CT series × 8 parameter variations × 6 ROIs described by 86 radiomics features. Intra-class variations are represented by colour shade variations. Green: normal liver tissue, blue: cysts, yellow: hemangioma, red: metastasis.

List of Supplemental Digital Content

Supplemental Digital Content 1 .docx

Figure 1

Click here to access/download;Figure;Fig_1.tiff

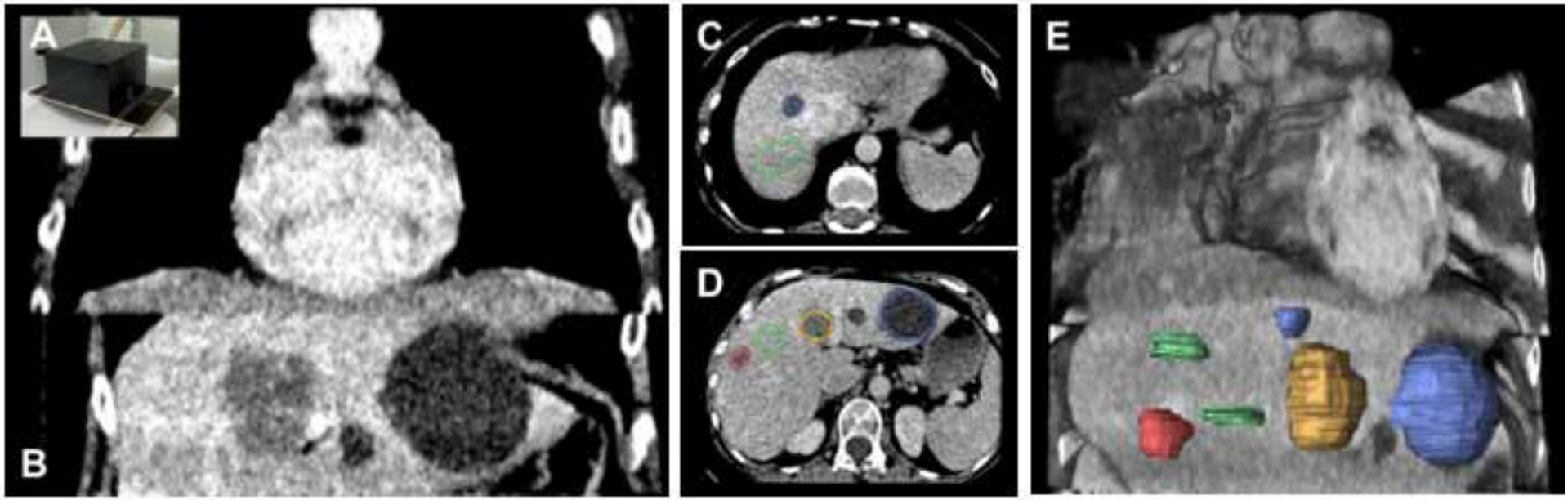


Figure 2

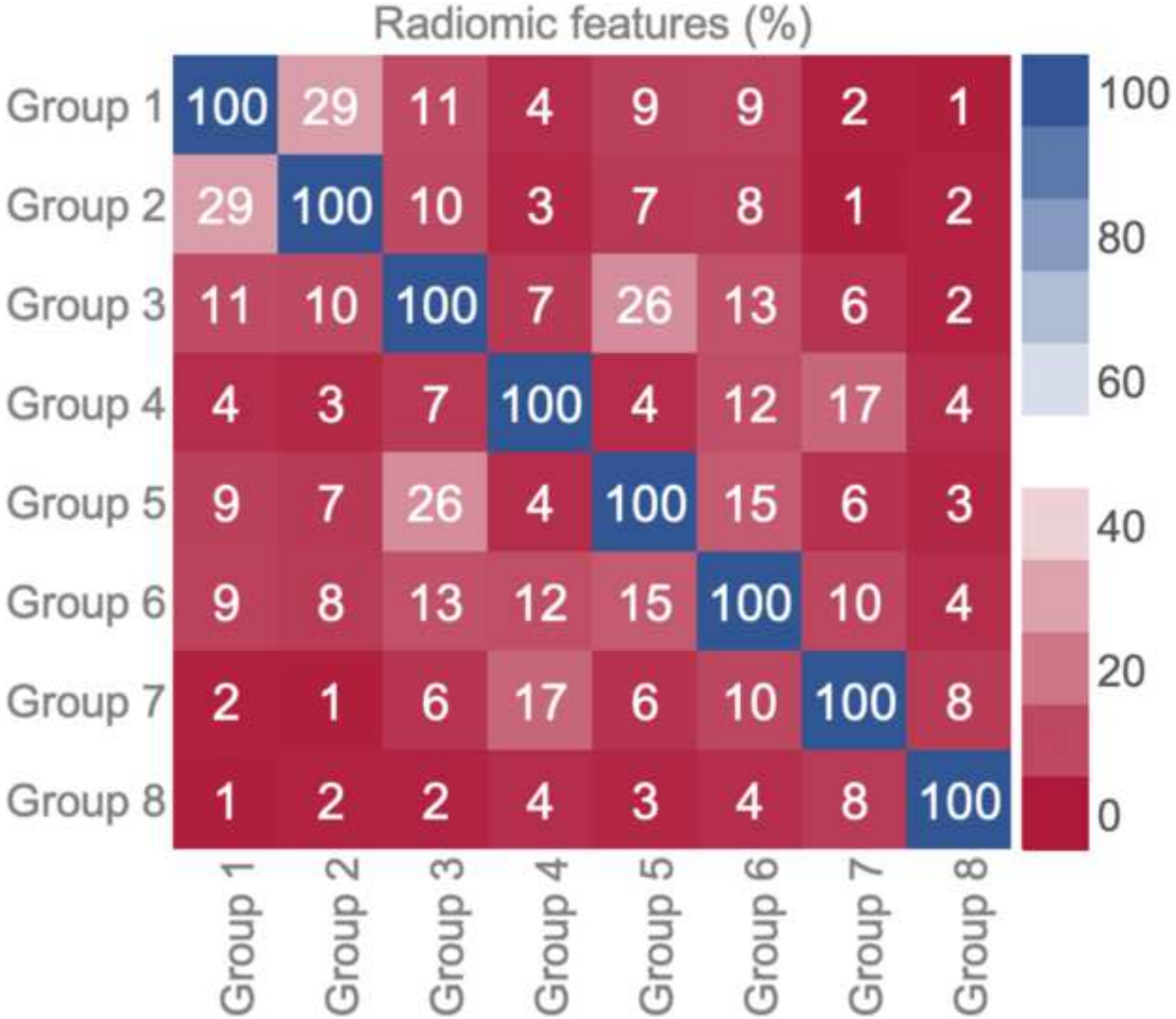


Figure 3

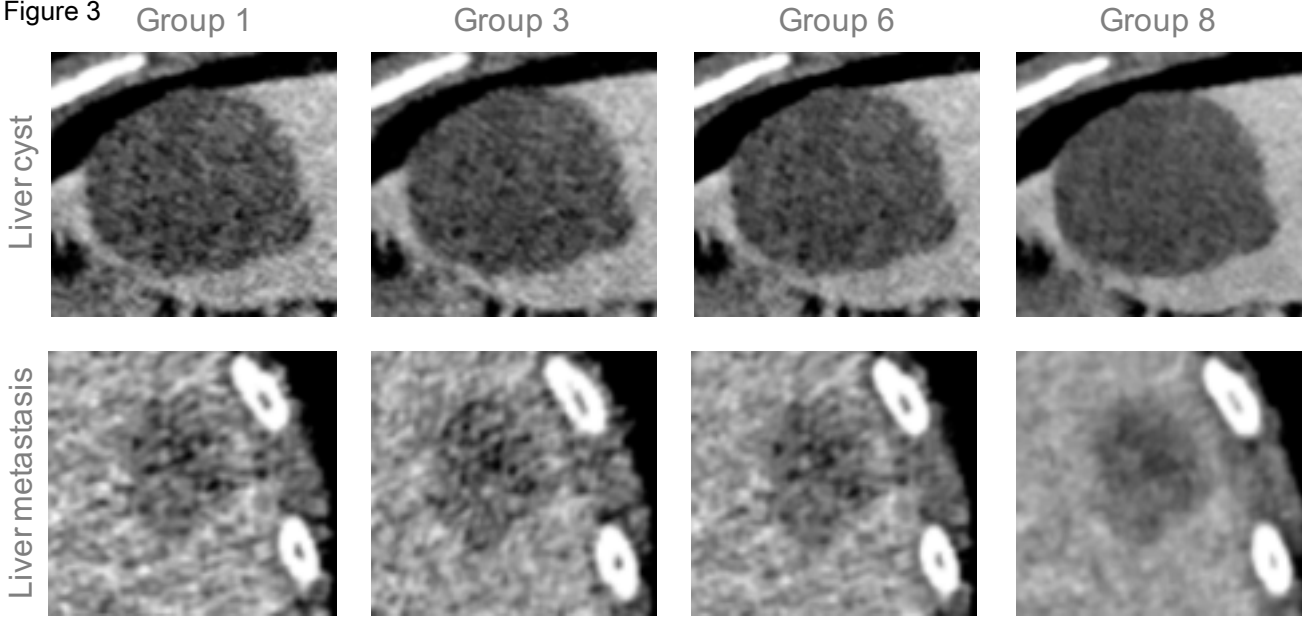


Figure 4

Click here to access/download;Figure;Fig_4.tiff

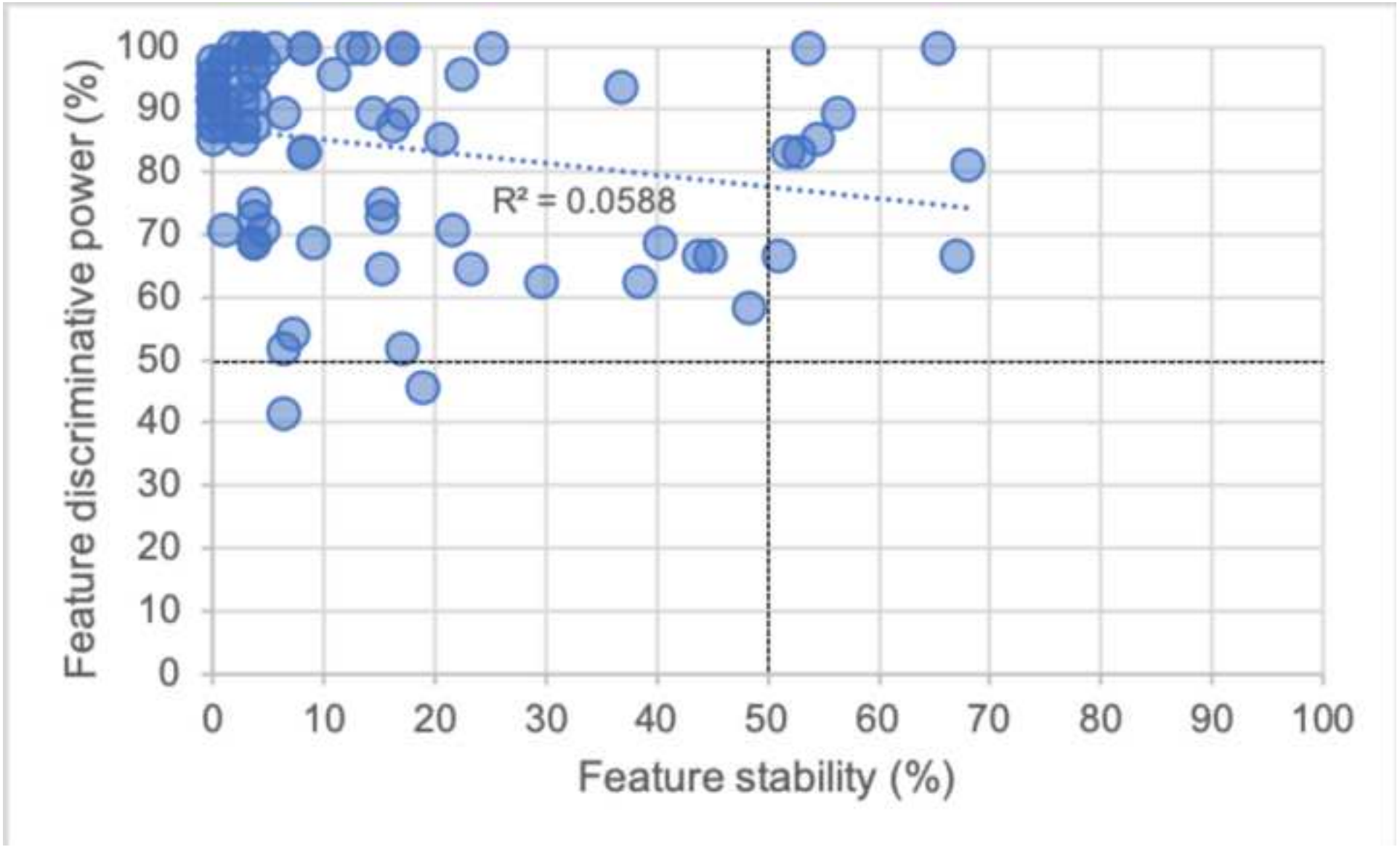
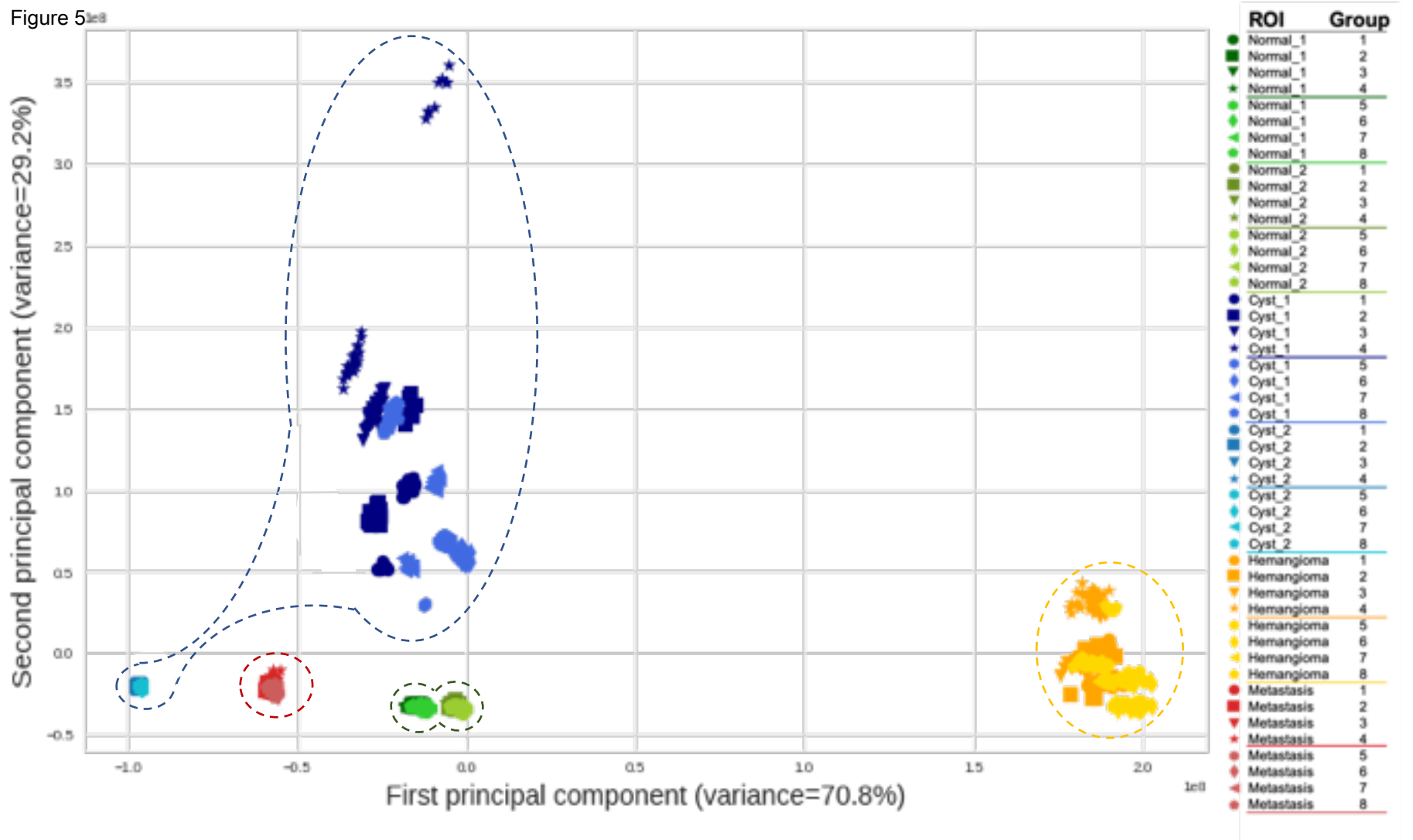


Figure 5


1e8
35
30
25
20
15
10
0.5
0.0
-0.5
Second principal component (variance=29.2%)
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
1e8
First principal component (variance=70.8%)
ROI
Group
Normal_1 1
Normal_1 2
Normal_1 3
Normal_1 4
Normal_1 5
Normal_1 6
Normal_1 7
Normal_1 8
Normal_2 1
Normal_2 2
Normal_2 3
Normal_2 4
Normal_2 5
Normal_2 6
Normal_2 7
Normal_2 8
Cyst_1 1
Cyst_1 2
Cyst_1 3
Cyst_1 4
Cyst_1 5
Cyst_1 6
Cyst_1 7
Cyst_1 8
Cyst_2 1
Cyst_2 2
Cyst_2 3
Cyst_2 4
Cyst_2 5
Cyst_2 6
Cyst_2 7
Cyst_2 8
Hemangioma 1
Hemangioma 2
Hemangioma 3
Hemangioma 4
Hemangioma 5
Hemangioma 6
Hemangioma 7
Hemangioma 8
Metastasis 1
Metastasis 2
Metastasis 3
Metastasis 4
Metastasis 5
Metastasis 6
Metastasis 7
Metastasis 8