08:00-09:00 **Research Stage 1**

Research Presentation Session: Artificial Intelligence & Machine Learning & Imaging Informatics

RPS 1705

Radiomics applications in MDCT

Moderator
R. Mirón Mombiela; Herlev/DK
(mirona@ufm.edu)

RPS 1705-2

A reference framework for standardisation and harmonisation of CT radiomics features: the "CadAIver" analysis

R. Levi[1], M. Mollura[2], G. Savini[1], F. Garoli[1], M. Battaglia[1], A. Ammirabile[1], L. A. Cappellini[1], R. Barbieri[2], L. S. Politi[1]; [1]Pieve Emanuele/IT, [2]Milan/IT
(riccardo.levi@st.hunimed.eu)

Purpose: The purpose of this study was to quantify the effects of different dose protocols, reconstruction algorithms, fields of view (FOVs) and CT scanners on radiomics features (RFs) of the lumbar vertebrae in a cadaveric trunk and to develop a normalisation algorithm to harmonise radiomics analyses.
Methods or Backround: We performed a total of 112 CT acquisitions of a cadaveric trunk on 3 different CT scanners from 2 different vendors, using varying kV (80-140) and mA (250-400). Each acquisition was performed using 2 FOVs (Abdomen/Spine) and two reconstruction kernels (Standard/Bone). Lumbar vertebrae were segmented using a convolutional neural network and RFs were extracted using pyradiomics. Intra/Inter-scanner analyses were assessed and each RF was tested using a generalised linear model (GLM) to assess the effects of all the above-mentioned CT acquisition parameters. GLM model was employed to standardise RFs across different acquisitions, and was compared to the ComBat algorithm, using 10-folds cross-validation (CV) R2.
Results or Findings: KV variation showed the highest feature modification in intra/inter-scanner analyses, with the FirstOrder features showing high variability (94.4%, Scanner 1). Little effects were evident upon mA variation. 100% of Shape features on all scanners were found significantly dependent on FOV, and 83% of GLSZM were statistically different between reconstruction kernels. The proposed GLM algorithm obtained a mean R2 across CV higher than 0.90 in 21 Radiomics features (19.6%), whereas ComBat obtained an high R2 value in 1 Radiomics feature (0.90%). GLM was statistically superior in 39 Radiomics features in respect to ComBat.
Conclusion: This study is the first attempt in describing the effects of CT acquisition parameters on RFs from cadaveric donor. Current and Voltage affect the RFs in different ways. GLM was superior to ComBat in normalising RFs across all different CT acquisitions.
Limitations: This study only featured vertebral analysis.
Ethics committee approval: Ethics committee approval is not required for the use of cadaveric body parts. The cadaver was donated willingly to science and obtained from medcure.org (Medcure, Orlando, FL, USA).
Funding for this study: No funding was received for this study.
Author Disclosures:
Riccardo Barbieri: Nothing to disclose
Luca Alessandro Cappellini: Nothing to disclose
Giovanni Savini: Nothing to disclose
Angela Ammirabile: Nothing to disclose
Letterio Salvatore Politi: Nothing to disclose
Federico Garoli: Nothing to disclose
Massimiliano Battaglia: Nothing to disclose
Maximiliano Mollura: Nothing to disclose
Riccardo Levi: Nothing to disclose

RPS 1705-3

Development of a radiomic model to predict the risk of hepatocellular carcinoma in cirrhotic patients

L. Leo, S. Schirò, O. Bazzini, F. Cazzato, D. Stefanelli, M. V. Bazzocchi, G. Milanese, G. Missale, N. Sverzellati; Parma/IT
(ludovica.leo@unipr.it)

Purpose: Patients with cirrhosis are at increased risk of hepatocellular carcinoma (HCC); clinical and radiological scores for predicting the risk of developing HCC are currently unknown. The aim of the study is to assess if radiomics can identify cirrhotic patients at risk of HCC.
Methods or Backround: 98 subjects (M:F = 64:34; mean age 67 years ±9.81) were included in this retrospective monocentric study. Two groups were identified: group (a) with 49 patients (M:F=31:18; mean age 67.4 years ±9.12) who had a baseline CT with radiological signs of cirrhosis and follow-up CTs without evidence of HCC (LI-RADS 1-3); group (b) with 49 patients (M:F = 33:16; mean age 67.5 years ±10.5) who had a baseline CT with radiological signs of cirrhosis and evidence of HCC in one follow-up CT scan (LI-RADS 4/5). Four radiologists (three years of experience) provided complete liver segmentations by manually drawing volumes of interest (VOI) on non-enhanced baseline CT scans, extracting 851 radiomic features (RF). 100 train:test (0.7:0.3) splits were created and recursive feature elimination with a 5-fold cross-validation was performed on train partitions using the random forest classifier (RFC). Subsequently, RFC was trained by selecting iteratively an increasing number of features sorted by their occurrences to evaluate the minimum number of informative features. Finally, means and 95% confidence intervals of accuracy, sensitivity, specificity, precision, area under the receiver operating characteristic curve (ROC-AUC) were calculated on the test partitions.
Results or Findings: The best model exploited two features (wavelet-LLH_glcm_DifferenceAverage, wavelet-HLH_gldm_DependenceVariance), reaching accuracy, sensitivity, specificity, precision and ROC-AUC of 0.72 [0.70-0.73], 0.76 [0.74-0.79], 0.68 [0.65-0.70], 0.71 [0.69-0.73] and 0.78 [0.76-0.79], respectively.
Conclusion: Our radiomic model, evaluated in cross-validation, predicted the development of HCC with satisfactory performances and was more sensitive than specific.
Limitations: There was a monocentric retrospective cohort and no external validation.
Ethics committee approval: This study was approved by an ethics committee; 930/2022/OS/UNIPR (Protocol 5853 07/02/2023).
Funding for this study: No funding was received for this study.
Author Disclosures:
Francesca Cazzato: Nothing to disclose
Ludovica Leo: Nothing to disclose
Daniela Stefanelli: Nothing to disclose
Nicola Sverzellati: Nothing to disclose
Gabriele Missale: Nothing to disclose
Maria Vittoria Bazzocchi: Nothing to disclose
Silvia Schirò: Nothing to disclose
Gianluca Milanese: Nothing to disclose
Olimpia Bazzini: Nothing to disclose

RPS 1705-4

Influence of CT scanners on radiomics features in abdominal CT: a multicentre phantom study

C. Aberle[1], M. Bach[1], O. Jimenez-del-Toro[2], R. Schaer[2], K. Flouris[3], E. Konukoglu[3], H. Müller[2], A. Depeursinge[2], *M. Obmann*[1]; [1]Basel/CH, [2]Sierre/CH, [3]Zurich/CH

Purpose: The aim of this study was to investigate the influence of different CT scanners on the stability and discriminative power of radiomics features using an anatomically accurate 3D-printed abdominal phantom.
Methods or Backround: Based on a patient's CT scan with multiple hepatic lesions, an anatomically and texturally realistic phantom was commercially 3D-printed using potassium-iodide ink on paper. The phantom was scanned on 13 CT scanners by 4 different manufacturers at 8 institutions with 10 scan repetitions each. A harmonised clinical oncologic CT acquisition protocol was used on all scanners. Images were reconstructed using iterative reconstruction algorithms. 86 radiomics features were assessed for six different ROIs (metastasis, haemangioma, 2 cysts, 2 normal liver parenchyma regions) using principal component analyses (PCA) and Kruskal-Wallis tests.
Results or Findings: For all ROIs, PCA analyses clearly showed clustering by scanners and manufacturers, with the same scanner models overlapping. Kruskal-Wallis tests for each ROI and radiomics feature showed significant differences between scanners in 511 of 516 tests ($P < 0.05$). Pairwise ROI comparison in the PCA showed both separation of the 13 different CT scanners and of the ROIs, while the separation between ROIs was stronger than between scanners.
Conclusion: In this multicentre study, radiomics features are impacted by CT scanner models in varying degrees, despite the use of matched acquisition and reconstruction parameters. When performing multicentre studies, an a priori phantom analysis and feature harmonisation techniques may be used to account for these influences and select more stable radiomics features.
Limitations: As the phantom includes 1-2 ROIs per tissue type, variability of the same tissue type was not studied and results of ROI separation may not be fully generalisable to tissue type classification. Patient motion cannot be assessed with this phantom and may aggravate interscanner variations.
Ethics committee approval: This was a phantom study.
Funding for this study: This work was partly supported by the Swiss Personalised Health Network with the QA4IQI Quality assessment for interoperable quantitative computed tomography imaging project DMS2445

and the IMAGINE project. It was also partially supported by the Swiss National Science Foundation (grant 205320_179069).
Author Disclosures:
Roger Schaer: Nothing to disclose
Markus Obmann: Nothing to disclose
Ender Konukoglu: Nothing to disclose
Michael Bach: Nothing to disclose
Adrien Depeursinge: Nothing to disclose
Kyriakos Flouris: Nothing to disclose
Christoph Aberle: Nothing to disclose
Henning Müller: Nothing to disclose
Oscar Jimenez-del-Toro: Nothing to disclose

RPS 1705-5

Radiomics and machine learning for the assessment of renal tumour histological subtypes on multiphase computed tomography: a multicentre trial with independent testing

S. Bachanek[*1], A. Leha[1], P. Zeuschner[2], A. Massmann[3], L. Trojan[1], J. Uhlig[1], J. Lotz[1], A. Uhlig[1]; [1]Göttingen/DE, [2]Homburg/DE, [3]Stuttgart/DE
(sophie.bachanek@med.uni-goettingen.de)

Purpose: The purpose of this study was to distinguish histological subtypes of renal tumours identified on multiphase computed tomography (CT) using radiomic features and machine learning in a multicentre setting.
Methods or Backround: Patients undergoing surgical resection and histopathological assessment of renal tumours at two tertiary urological centres between 2012 and 2022 were retrospectively included. Preoperative arterial and venous phase CTs from multiple referring imaging centres were segmented and standardised radiomic features extracted. After preprocessing and class imbalance, an extreme gradient boosting tree-based (XGB) machine learning (ML) algorithm was used to predict renal tumour subtypes using 10-fold cross-validation, assessed as multiclass AUC. ML algorithms were trained on data from one centre, and independently tested on data from the other centre.
Results or Findings: The training cohort comprised n=297 patients (n=191 ccRCC, n=40 pRCC, n=22 chRCC, n=28 oncocytomas, and n=16 AML), and the testing cohort n=121 patients (n=68/ n=20/ n=4/ n=26/ n=3). The XGB algorithm demonstrated a diagnostic performance of AUC=0.81/ 0.64/ 0.8 for venous / arterial / combined contrast phase CT in the training cohort, and AUC=0.75/ 0.67/ 0.75 in the independent validation cohort. In pairwise comparisons, the lowest diagnostic accuracy was evident for identification of oncocytomas (AUC=0.57-0.69), and the highest for identification of AMLs (AUC=0.9-0.94).
Conclusion: Radiomic feature analyses acquired from clinical routine CT yield robust results for renal tumour assessment. For renal tumour subtype discrimination, venous phase CT yields the most pertinent imaging information, without evident diagnostic benefit of an added arterial contrast phase. Among all renal tumours, oncocytomas are hardest to differentiate using CT.
Limitations: Limitations of the study mainly include the heterogeneity of renal tumour subtypes in the training and testing cohort with occasional resulting low case numbers in the subgroups.
Ethics committee approval: The study received ethical approval by the ethics committee at Goettingen University (No. 2/4/17) and at Saarland University (No. 67/19).
Funding for this study: No funding was received for this study.
Author Disclosures:
Annemarie Uhlig: Nothing to disclose
Lutz Trojan: Nothing to disclose
Andreas Leha: Nothing to disclose
Sophie Bachanek: Nothing to disclose
Johannes Uhlig: Nothing to disclose
Joachim Lotz: Nothing to disclose
Alexander Massmann: Nothing to disclose
Philip Zeuschner: Nothing to disclose

RPS 1705-6

Overcoming data scarcity in radiomics/radiogenomics using synthetic radiomic features

M. Ahmadian, Z. Bodalal, H. J. van der Hulst, C. Vens, L. Karssemakers, R. G. H. Beets-Tan, M. van den Brekel, J. A. Castelijns; Amsterdam/NL
(m.ahmadian@nki.nl)

Purpose: This study evaluates the potential of synthetic radiomic data generation to augment the performance of radiomics/radiogenomics prediction models.
Methods or Backround: This study was conducted on a retrospectively collected cohort of 386 colorectal cancer patients for whom matched contrast-enhanced CT images and TP53 mutational status were available. Five different tabular synthetic data generation models were used to generate synthetic radiomic data based on real-world radiomics extracted from our cohort. The quality and reproducibility of the generated synthetic radiomic data were assessed. Synthetic radiomics were then combined with real-world radiomic training data to evaluate their impact on the predictive model's performance.
Results or Findings: Using only real-world radiomic data, increasing training samples (n=200, 400, 1000, 2055 lesions) improved the model's predictive performance on the unseen test set (average AUC=0.52, 0.53, 0.56, and 0.64). Synthetic tabular data generation models created reproducible synthetic radiomic data with properties highly similar to real-world data (for n =1000 lesions, average Chi-square =0.932, average basic statistical correlation =0.844). Augmenting predictive models with synthetic radiomics enhanced their performance by 9.61%, 11.32%, 16.07%, and 3.22% for the outlined training sets, respectively. Synthetic radiomics derived from randomised/noisy radiomic data failed to enhance predictive performance, while true signal data was effectively amplified.
Conclusion: Synthetic radiomic data, when combined with real radiomics, can enhance the performance of predictive models. Tabular synthetic data generation might help overcome limitations in medical AI stemming from data scarcity.
Limitations: Our study was limited to a retrospective monocentric cohort. It would be beneficial to validate our findings on external cohorts, helping to ensure that the conclusions were not specific to the centre where the data was gathered.
Ethics committee approval: This study was approved by the Institutional Review Board (IRB): IRBd19-147.
Funding for this study: This work was made possible via a grant from the Hanarth Foundation. The computational infrastructure for the analysis was funded via the Maurits en Anna de Kock Stichting and the NVIDIA Academic GPU programme.
Author Disclosures:
Zuhir Bodalal: Nothing to disclose
Conchita Vens: Nothing to disclose
Milad Ahmadian: Nothing to disclose
Luc Karssemakers: Nothing to disclose
Jonas A. Castelijns: Nothing to disclose
Regina G. H. Beets-Tan: Nothing to disclose
Hedda Joanne van der Hulst: Nothing to disclose
Michiel van den Brekel: Nothing to disclose

RPS 1705-7

Impact of population size and validation method on the performance of radiomics models: application to COVID lung lesions

A. Decoux, L. Duron, A. Arnoux, L. S. Fournier; Paris/FR

Purpose: The purpose of this study was to explore the impact of population size and validation strategy on the estimated performance and reproducibility of radiomics studies.
Methods or Backround: Radiomics parameters were extracted from lung lesions segmented by experts on CT in 3,737 COVID-19 patients (STOIC cohort). 1121 (33%) patients were set aside to simulate an external validation population ("generalisation set"). Among the remaining patients, subpopulations of varying sizes were generated to simulate the training/test population for the radiomics study. Prediction models were trained on 100 bootstrapped samples to estimate variance of the AUC, i.e. model stability. Three validation strategies were tested: one time split, cross-validation and nested cross-validation. The mean and variance of AUCs of each model was calculated on the subpopulation as well as on the "generalisation" set, and the difference was the generalisation gap.
Results or Findings: Increasing the size of the training data sets improved model performance on both internal validation and generalisation sets, decreased the variance of performance on the validation set and decreased the generalisation bias, thereby increasing overall confidence in the model, with a plateau at 400 patients. Cross-validation helped reduce variance and generalisation bias compared to one time split. Nested cross-validation reduced variance but at the expense of increased generalisation bias.
Conclusion: As expected, population size has a strong impact on model performance, particularly on the estimated performance variance (stability) of models. This study is the first to estimate the minimum population size needed to improve generalisability of radiomics studies. However, as it is applied to a single data set, results are expected to vary according to predictive power of imaging for a given clinical question.
Limitations: Our generalisation set serves as a surrogate for an external validation set, it doesn't constitute a true external validation set.
Ethics committee approval: This is a retrospective study.
Funding for this study: This work was funded by the French government under management of the Agence Nationale de la Recherche as part of the "Investissements d'avenir" programme, reference ANR19-P3IA-0001 (PRAIRIE 3IA Institute).
Author Disclosures:
Laure S. Fournier: Nothing to disclose
Loïc Duron: Nothing to disclose
Armelle Arnoux: Nothing to disclose
Antoine Decoux: Nothing to disclose