



OPEN

Assessing robustness of carotid artery CT angiography radiomics in the identification of culprit lesions in cerebrovascular events

Elizabeth P. V. Le¹, Leonardo Rundo^{2,3}, Jason M. Tarkin¹, Nicholas R. Evans^{1,4}, Mohammed M. Chowdhury⁵, Patrick A. Coughlin⁵, Holly Pavey⁶, Chris Wall¹, Fulvio Zaccagna^{2,7}, Ferdia A. Gallagher², Yuan Huang^{2,8}, Rouchelle Sriranjani¹, Anthony Le⁹, Jonathan R. Weir-McCall², Michael Roberts^{8,10,11}, Fiona J. Gilbert², Elizabeth A. Warburton⁴, Carola-Bibiane Schönlieb^{8,11}, Evis Sala^{2,3} & James H. F. Rudd^{1✉}

Radiomics, quantitative feature extraction from radiological images, can improve disease diagnosis and prognostication. However, radiomic features are susceptible to image acquisition and segmentation variability. Ideally, only features robust to these variations would be incorporated into predictive models, for good generalisability. We extracted 93 radiomic features from carotid artery computed tomography angiograms of 41 patients with cerebrovascular events. We tested feature robustness to region-of-interest perturbations, image pre-processing settings and quantisation methods using both single- and multi-slice approaches. We assessed the ability of the most robust features to identify culprit and non-culprit arteries using several machine learning algorithms and report the average area under the curve (AUC) from five-fold cross validation. Multi-slice features were superior to single for producing robust radiomic features (67 vs. 61). The optimal image quantisation method used bin widths of 25 or 30. Incorporating our top 10 non-redundant robust radiomics features into ElasticNet achieved an AUC of 0.73 and accuracy of 69% (compared to carotid calcification alone [AUC: 0.44, accuracy: 46%]). Our results provide key information for introducing carotid CT radiomics into clinical practice. If validated prospectively, our robust carotid radiomic set could improve stroke prediction and target therapies to those at highest risk.

Carotid CT angiography (CTA) is commonly performed following an ischaemic stroke or transient ischaemic attack (TIA) to help guide patient management, for example between carotid endarterectomy surgery plus medical therapy or medical therapy alone. Carotid CTA imaging allows measurement of carotid artery luminal stenosis and unenhanced images provide information about calcification of the artery wall. The decision to perform carotid endarterectomy surgery, to reduce future stroke risk, is based on the degree of carotid stenosis and the presence of relevant symptoms. However, whilst stenosis provides important information about disease burden, it does not inform about the underlying plaque stability or degree of inflammation and patients may have second events with only mild to moderate carotid artery narrowings¹.

Radiomics, sometimes called 'texture analysis', comprises image analysis methods that involve the high-throughput extraction of minable imaging features² from radiological images. Radiomic features quantify simple and complex patterns in the data, such as the roundness of a tumour, the spatial arrangement of voxels or variations in signal intensity across a lesion of interest. These features have been used to develop diagnostic and

¹Department of Medicine, University of Cambridge, Cambridge, UK. ²Department of Radiology, University of Cambridge, Cambridge, UK. ³Cancer Research UK Cambridge Centre, University of Cambridge, Cambridge, UK. ⁴Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK. ⁵Division of Vascular Surgery, Department of Surgery, University of Cambridge, Cambridge, UK. ⁶Division of Experimental Medicine and Immunotherapeutics, University of Cambridge, Cambridge, UK. ⁷Department of Medical Imaging, University of Toronto, Toronto, ON, Canada. ⁸EPSRC Centre for Mathematical Imaging in Healthcare, University of Cambridge, Cambridge, UK. ⁹School of Medicine, University of Leeds, Leeds, UK. ¹⁰Oncology R&D, AstraZeneca, Cambridge, UK. ¹¹Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK. ✉email: jhfr2@cam.ac.uk

prognostic prediction models, particularly in oncology^{3,4} where radiomic features can predict lung and other cancer survival times better than the gold standard TNM (Tumour, Node, Metastasis) staging system^{5–7}. There is growing interest in the application of radiomics to cardiovascular imaging^{8–11}, for example to differentiate causes of prosthetic valve obstruction using cardiac computed tomography (CT) radiomic features⁹, to distinguish between hypertensive and hypertrophic cardiomyopathy from magnetic resonance imaging (MRI)¹⁰, or to characterise carotid artery plaques from ultrasound images^{12,13}.

A radiomic biomarker should be reproducible, robust and accurate¹⁴. However, radiomic features are susceptible to variations^{15,16}, including image acquisition (e.g. use of different CT scanner manufacturers and models, acquisition protocols and image reconstruction methods), image segmentation (e.g. inter-observer and intra-observer variability in delineating the region-of-interest [ROI]/volume-of-interest [VOI]) and at the feature extraction stage (e.g. use of different radiomics software, different image pre-processing settings or radiomic feature definitions). To minimise such variations⁶, there is a growing call for the standardisation of protocols at every stage of the radiomics workflow¹⁷. Where harmonisation is not possible (e.g. using the same type of CT scanner in every hospital), robustness analyses are essential in determining the extent to which such variations can be tolerated for each specific application, i.e. without affecting predictive performance. Robustness analyses evaluate the impact of changes in these parameters on radiomic features, aiming to find those most immune to such perturbations. These ‘robust’ features are expected to perform well when tested on new image datasets, a characteristic referred to as ‘good generalisability’¹⁸.

The majority of published robustness and repeatability studies have been conducted using phantoms^{19,20} and restricted to oncology^{21–23}, such as in non-small cell lung cancer¹⁶ or oesophageal cancer²⁴. In cardiovascular imaging, there have been relatively few studies—one using a phantom in single photon emission computed tomography²⁵ and the other finding robust myocardial radiomic features from cardiac MRI²⁶. However, feature robustness is specific to the disease phenotype being studied and to the imaging modality used¹⁸. Therefore, there is an unmet need for the assessment of radiomic robustness in cardiovascular disease, specifically in carotid CT angiography (CTA) imaging.

In this study, we first investigated the robustness of 93 individual carotid CTA radiomic features following ROI/VOI perturbations under different CTA image pre-processing and across single vs multiple artery slice situations. We then determined (1) the optimal image pre-processing settings (i.e. the settings that provided the highest proportion of radiomic features with excellent robustness) and (2) the most robust and non-redundant (i.e. not highly correlated) radiomic features for machine learning classification of culprit *versus* non-culprit carotid arteries in patients with prior cerebrovascular events.

In summary, we sought to understand whether radiomic features extracted from standard clinical CT scans were robust and reliable and whether they could provide additional prognostic information to help identify higher-risk culprit arteries from lower-risk non-culprit carotid arteries.

Results

Carotid CTA scans from 41 patients with previous stroke or TIA were analysed in this study comprising 41 culprit and 41 non-culprit carotid arteries (82 carotid arteries in total). The clinical characteristics of the patients and the plaque characteristics of their carotid arteries are shown in Table 1.

Assessing the ability of morphological operations to capture inter-observer segmentation variability

We found that there was low variability in intra-observer segmentation, as shown in Fig. 1, but there was greater variability in inter-observer segmentation. The morphological operations applied to the ROIs captured the range of variability that occurred with human inter-observer variability, demonstrated in Fig. 1.

Assessing feature robustness in different image configurations

Feature robustness in single-slice analysis. Over 50% of radiomic features that were extracted had excellent robustness to ROI perturbations when using the original image (i.e. no prior normalisation or resegmentation) and a fixed BW for image quantisation, ranging from 10 to 35 (in increments of 5). Using a fixed BW of 10 led to a higher proportion of poorly robust (ICC < 0.5) radiomic features compared with the other BWs, see Supplementary Fig. S1.

The best BW setting (i.e. had the most features with excellent robustness) for single-slice analysis with no image pre-processing was BW 25–30. This corresponds with the PyRadiomics default setting (fixed BW of 25) and a detailed breakdown of the radiomic features by robustness category and feature class type for this setting is provided in Supplementary Table S1. Overall, using a fixed BW of 25 in single-slice analysis resulted in 52.7% of radiomics features having excellent robustness, 35.5% having moderate robustness and 11.8% having poor robustness. Using a fixed BW rather than a fixed BN (from 8 to 256) for image quantisation led to a higher proportion of radiomic features with excellent robustness. If using fixed BNs as the method for image quantisation, the upper limit of BNs investigated (from 8 to 256 in powers of 2) led to the highest proportion of poorly robust radiomic features (15.1% of radiomic features had poor robustness, whilst the proportions were < 10% for other BNs).

Prior normalisation of the image reduced the proportion of poorly robust radiomic features but did not impact the proportion of radiomic features with excellent robustness, when compared to no prior normalisation. In the prior normalisation pre-processing setting, only different BNs could be investigated as using different BWs in the range of 10–35 led to ROIs with too few grey values for radiomic feature calculation.

Resegmentation in single-slice analysis reduced the proportion of radiomic features with excellent robustness (from 52.7 to 8.6%), but also reduced the proportion of poorly robust radiomic features as compared to no

Patient characteristics (n = 41)			
Age (years)—mean (SD)	74.1 (8.4)		
Male—n (%)	32 (78.0%)		
<i>Presence of cardiovascular risk factors</i>			
Diabetes mellitus—n (%)	8 (19.5%)		
Hypertension—n (%)	27 (65.9%)		
Smoking history (current or former)—n (%)	29 (70.7%)		
<i>Cerebrovascular event</i>			
Stroke—n (%)	30 (73.2%)		
TIA—n (%)	11 (26.8%)		
Carotid characteristic	Culprit (n = 41)	Non-culprit (n = 41)	p value
<i>Plaque type</i>			
Calcified—n (%)	9 (22.0%)	15 (36.6%)	0.070
Noncalcified—n (%)	2 (4.9%)	5 (12.2%)	0.250
Mixed—n (%)	30 (73.2%)	21 (51.2%)	0.012*
<i>Calcium burden (AU)</i>			
Median (IQR)	263 (95–701)	387 (63–659)	0.706
Minimum	0	0	-
Maximum	1963	1671	-
<i>Carotid stenosis (%)</i>			
Mean (SD)	72 (17)	40 (22)	4.70 × 10 ⁻⁹ ***
Minimum	29	3	-
Maximum	99	88	-

Table 1. Patient and culprit versus non-culprit carotid characteristics. Plaque type comparisons: McNemar's test on paired proportions; Calcium score comparison: Wilcoxon signed-rank test; Carotid stenosis comparison: paired Student t-test. n, number of carotids or number of patients respectively; AU, Agatston units; IQR, interquartile range, SD, standard deviation; TIA, transient ischaemic attack; * p value < 0.05, *** p value < 0.001.

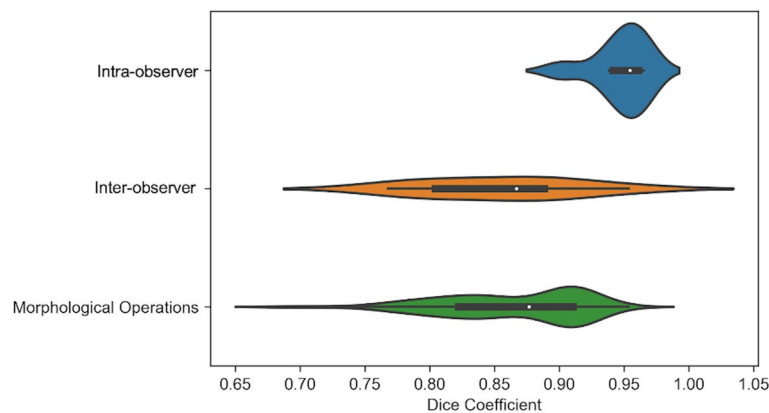


Figure 1. Violin plots of image segmentation agreement as determined by the Dice coefficient with the original ROI. Intra-observer variability was determined by manual segmentation of the ROIs by the same observer (EPVL) at two separate time points for 8 carotid arteries. Inter-observer variability was determined by manual ROI segmentation by two independent observers (EPVL and CW), performed on 8 carotid arteries. The ROIs drawn by the primary observer (EPVL) on 82 carotid arteries were compared with those generated following dilation and erosion morphological operations in single-slice analysis to determine the Dice coefficient distribution of systematic ROI perturbations. The violin plots each contain a black box-plot of the Dice coefficients, with the white dot representing the median Dice coefficient. Whereas the box-plots correspond to the actual data points of the Dice coefficients, the coloured shapes provide a visualisation of the underlying Dice coefficient distributions via kernel density estimations.

Feature class	Excellent robustness ICC ≥ 0.9	Moderate robustness 0.5 ≤ ICC < 0.9	Poor robustness ICC < 0.5
Multi-slice Analysis: Original, fixed bin width = 25, B-spline interpolation			
First order (n = 18)	10 (55.6%)	6 (33.3%)	2 (11.1%)
GLCM (n = 24)	17 (70.8%)	7 (29.2%)	0 (0.0%)
GLDM (n = 14)	5 (35.7%)	7 (50.0%)	2 (14.3%)
GLRLM (n = 16)	11 (68.8%)	1 (6.3%)	4 (25.0%)
GLSZM (n = 16)	6 (37.5%)	8 (50.0%)	2 (12.5%)
NGTDM (n = 5)	3 (60.0%)	2 (40.0%)	0 (0.0%)
Total (n = 93)	52 (55.9%)	31 (33.3%)	10 (10.8%)
Multi-slice Analysis: Resegmentation, fixed bin width = 25, B-spline interpolation			
First order (n = 18)	3 (16.7%)	13 (72.2%)	2 (11.1%)
GLCM (n = 24)	2 (8.3%)	22 (91.7%)	0 (0.0%)
GLDM (n = 14)	3 (21.4%)	11 (78.6%)	0 (0.0%)
GLRLM (n = 16)	3 (18.8%)	13 (81.3%)	0 (0.0%)
GLSZM (n = 16)	3 (18.8%)	13 (81.3%)	0 (0.0%)
NGTDM (n = 5)	0 (0.0%)	5 (100.0%)	0 (0.0%)
Total (n = 93)	14 (15.1%)	77 (82.8%)	2 (2.2%)

Table 2. Number of radiomic features with excellent, moderate and poor robustness by feature class in multi-slice analysis. Definition of ICC used = 2-way mixed-effects model, absolute agreement, single rater intraclass correlation coefficient; n, number of radiomics features.

resegmentation (from 11.8 to 3.2%). In resegmentation, the majority of radiomic features had moderate robustness against ROI perturbations, 88.2% compared with only 35.5% when no image pre-processing was applied.

Across the 19 different image settings investigated (see Supplementary Table S2), only 2 radiomic features (a) GLDM: Large Dependence High Grey Level Emphasis and (b) GLRLM: Long Run High Grey Level Emphasis, demonstrated excellent robustness across all 19 settings (100%). However, 61 out of 93 (65.6%) extracted radiomic features showed excellent robustness in at least 1 setting, see Supplementary Fig. S2.

Feature robustness in multi-slice analysis. In multi-slice analysis, the image quantisation method leading to the highest proportion of radiomic features with excellent robustness involved using fixed BWs, as opposed to fixed BNs, consistent with our single-slice analysis findings. Over 55% of all radiomic features had excellent robustness when using the original image in multi-slice analysis, with similar proportions across all BW settings from 10 to 35. There was a slight decrease in performance between BW 30 to BW 35 (BW 30; 58.1% excellent robustness, 9.7% poor robustness; BW 35; 55.9% excellent robustness, 11.8% poor robustness).

Similar to our single-slice analysis findings, prior normalisation in multi-slice analysis reduced the proportion of poorly robust features but had little impact on the proportion of excellent robustness radiomic features. Following resegmentation, the proportion of features with excellent robustness decreased (from 55.9% to 15.1%), however, so did the proportion of poorly robust features (from 10.8% to 2.2%). The majority of radiomic features shifted to moderate robustness, 82.8% compared with 33.3% when no image pre-processing was applied. Table 2 provides a breakdown of the different radiomic feature classes by robustness category (excellent, moderate and poor) in multi-slice analysis following (A) no image pre-processing and (B) resegmentation.

Across the 19 image settings investigated in multi-slice analysis (see Supplementary Table S2), 4 radiomic features demonstrated excellent robustness across all settings (100%), these were: a) GLDM: Grey Level Variance, b) First Order: Mean Absolute Deviation, c) GLRLM: Grey Level Variance and d) GLDM: Large Dependence High Grey Level Emphasis. In at least 1 setting, 67 out of 93 (72%) extracted radiomic features showed excellent robustness, see Supplementary Fig. S3.

Radiomic feature robustness similarities and differences in single-slice and multi-slice approaches. In single-slice analysis, 61 features had excellent robustness in at least one image setting out of the 19 settings investigated, whilst 67 features had excellent robustness in multi-slice analysis. There was considerable overlap in the radiomic features with excellent robustness between single-slice and multi-slice approaches (n = 56), these included First Order: Variance and GLCM: Autocorrelation. However, there were also radiomic features that had excellent robustness in the single-slice approach only (n = 5) such as First Order: Kurtosis and GLDM: Small Dependence Emphasis or the multi-slice approach only (n = 11) such as First Order: Uniformity and GLCM: Joint Energy, these are illustrated in Fig. 2A.

In single-slice analysis, 18 features were identified as poorly robust in at least 1 out of the 19 settings investigated, whilst 23 features were identified as poorly robust in multi-slice analysis. There were no radiomic features that were poorly robust in all 19 settings. Figure 2B illustrates the radiomic features that were identified as poorly robust in both single-slice and multi-slice analysis (n = 13) such as First Order: 10th Percentile and GLDM: Low Grey Level Emphasis, as well as those unique to single-slice (n = 5) analysis such as GLSZM: Grey Level Variance and NGTDM: Contrast or multi-slice analysis (n = 10) such as GLCM: Inverse Variance and NGTDM: Coarseness.

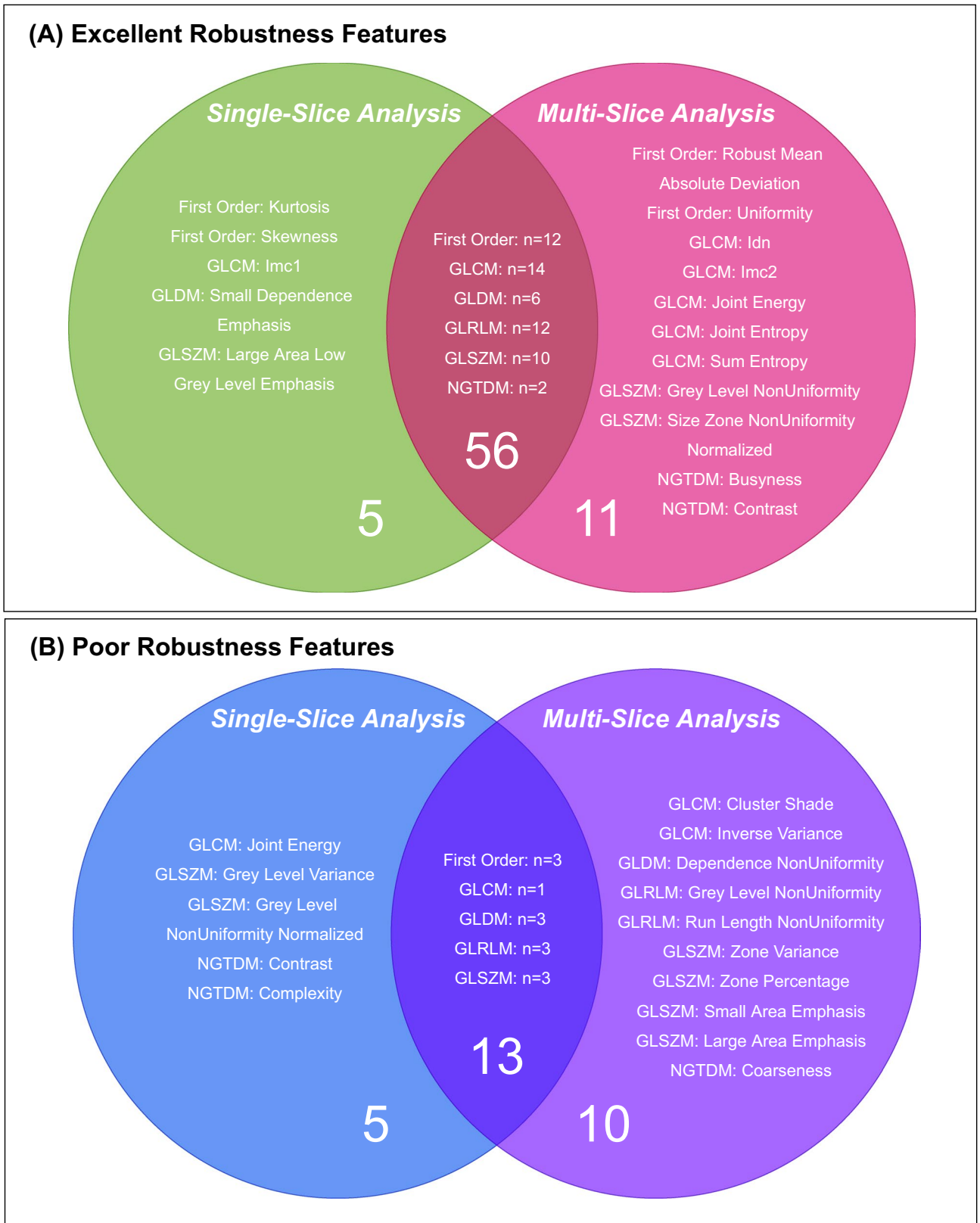


Figure 2. Venn diagram to show the radiomic features that had (A) Excellent robustness in single-slice and/or multi-slice analysis and, (B) Poor robustness in single-slice and/or multi-slice analysis.

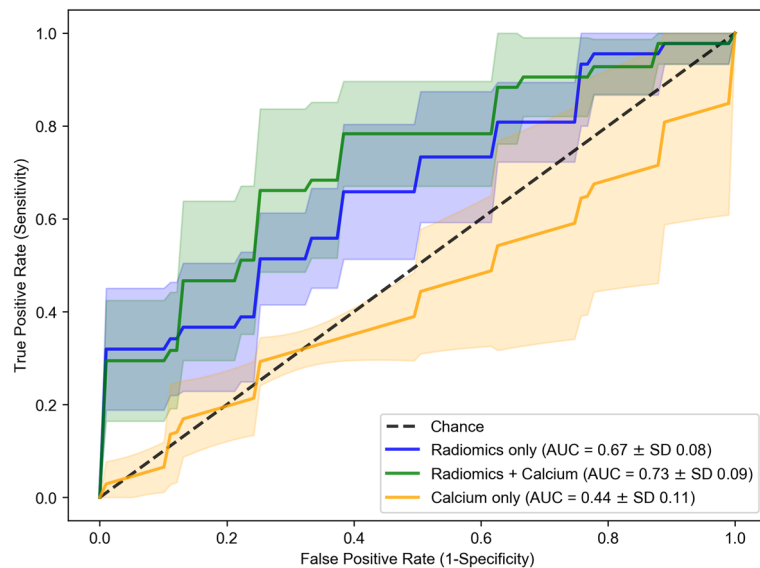


Figure 3. Mean receiver operating characteristic (ROC) curves of five-fold stratified cross-validation in multi-slice analysis with resegmentation for Elastic Net logistic regression (weight for L_1 and L_2 penalties = 0.5) using (1) radiomic features only as predictors, (2) radiomics features and calcium and (3) calcium only. Image setting: multi-slice analysis with resegmentation and a fixed bin width of 25. Dashed line indicates expected AUC for a random chance classifier. AUC, area under the ROC curve; SD, standard deviation.

Multi-slice analysis: Impact of PyRadiomics interpolation method

The proportion of robust radiomic features was similar whether B-spline or linear interpolation was used to resample the 3 mm slice thickness images and VOI segmentation masks to $1 \times 1 \times 1 \text{ mm}^3$. The vast majority of radiomic features showed excellent absolute agreement and consistency between the two methods of interpolation. No features were poorly robust with regards to the method of interpolation used. The breakdown by radiomic feature class is shown in Supplementary Table S3.

Culprit *versus* non-culprit carotid arteries: Machine learning classification

Since a fixed bin width of 25 was found to produce the highest proportion of radiomic features with excellent robustness, we used the radiomic features extracted using this image quantisation level in the following image settings for machine learning classification: (1) single-slice approach: original image, (2) single-slice approach: with resegmentation, (3) multi-slice approach: original image and (4) multi-slice approach: with resegmentation.

Non-redundant radiomic feature sets with excellent robustness in single-slice and multi-slice approaches. Different sets of non-redundant radiomic features with excellent robustness were identified depending on the image setting used (1–4), these radiomic features are detailed in Supplementary Table S4. For single-slice analysis using the original image, this consisted of 14 radiomic features; following resegmentation, this comprised 7 radiomic features. For multi-slice analysis using the original image, this consisted of 14 radiomic features that decreased to 10 radiomic features following resegmentation.

Machine learning classification performance determined by five-fold cross-validation. Several machine learning classifiers were investigated in a five-fold cross-validation scheme using (1) carotid calcium score as the only predictor, (2) radiomic features (non-redundant with excellent robustness) as the only predictors and (3) radiomic features with carotid calcium score (termed the ‘integrated model’) as predictors to differentiate culprit from non-culprit carotid arteries. The image setting that led to the highest predictive performance was the multi-slice approach with resegmentation (image setting 4). Within this setting, the best performing machine learning classifier amongst those investigated was the Elastic Net logistic regression-based classifier. Elastic Net regression uses a mixture between L_1 and L_2 regularisation whereby L_1 regularisation reduces the coefficients of certain features to zero, thereby reducing the number of variables in a model (i.e. sparse feature selection) and the L_2 penalty term constrains the magnitude of the feature coefficients so that a model is not dominated by any single feature. In this image setting and using this best performing model (Elastic Net, weight for L_1 and L_2 penalties = 0.5), carotid calcium score alone was a poor predictor of culprit *versus* non-culprit carotid artery status, see Supplementary Table S5.

The mean (standard deviation, SD) area under the receiver operating characteristic curve (AUC) for carotid calcium score alone was 0.44 (0.11) and the mean (95% confidence intervals [CI]) accuracy was 46% (25–56%), see Fig. 3. Please note, in five-fold cross-validation, an AUC is provided for the model performance in each fold. The mean cross-validated AUC is the average of the AUC values across the five folds.

		Cross Validation Fold					
		1	2	3	4	5	
		(A) Calcium Only					
AUC		0.37	0.5	0.55	0.5	0.25	Mean (SD): 0.44 (0.11)
		Feature Coefficients: (B) Radiomic Features Only					Feature Class
Radiomic Feature							
Dependence Variance		0.746	0.734	0.611	0.766	0.381	GLDM
Size Zone NonUniformity		0.029	0.283	0.798	0.426	0.541	GLSZM
Grey Level NonUniformity		-0.508	-0.930	-0.896	-0.395	-0.588	GLSZM
Long Run High Grey Level Emphasis		-1.077	-0.713	-0.859	-0.766	-0.993	GLRLM
Difference Variance		-0.960	0	-0.601	-0.580	-0.772	GLCM
Grey Level NonUniformity Normalized		-0.162	0	-0.151	-0.080	0	GLRLM
Large Area High Grey Level Emphasis		0	0	0	-0.034	0.017	GLSZM
Mean Absolute Deviation		-0.030	0	0	0	0.005	First Order
Large Dependence High Grey Level Emphasis		0	0.132	0	0	0	GLDM
Joint Energy		0.033	0	0	0	0	GLCM
AUC		0.59	0.56	0.73	0.77	0.69	Mean (SD): 0.67 (0.08)
p-value		0.074	0.690	0.247	0.033*	0.088	
		Feature Coefficients: (C) Radiomic Features & Calcium					Feature Class
Radiomic Feature							
Dependence Variance		0.864	0.999	0.812	0.847	0.573	GLDM
Calcium		0.396	0.687	0.648	0.478	0.642	Clinical
Grey Level NonUniformity		-0.508	-0.915	-0.750	-0.256	-0.479	GLSZM
Long Run High Grey Level Emphasis		-1.114	-0.769	-1.024	-0.881	-1.172	GLRLM
Size Zone NonUniformity		-0.033	0	0.328	0.054	0.072	GLSZM
Difference Variance		-1.008	0	-0.604	-0.612	-0.808	GLCM
Grey Level NonUniformity Normalized		0	0.061	-0.065	0	0	GLRLM
Large Area High Grey Level Emphasis		0	0	0	-0.009	0.022	GLSZM
Large Dependence High Grey Level Emphasis		0	0.070	0	0	0	GLDM
Joint Energy		0.055	0	0	0	0	GLCM
Mean Absolute Deviation		0	-0.080	0	0	0	First Order
AUC		0.63	0.63	0.78	0.84	0.75	Mean (SD): 0.73 (0.09)
p-value		0.054	0.403	0.065	9.83x10 ⁻⁴ ****	0.020*	

Figure 4. AUC and feature coefficients of predictors used in the Elastic Net logistic regression classifier in multi-slice analysis with resegmentation. The coefficients of the non-redundant radiomic features with excellent robustness in multi-slice analysis with resegmentation according to the Elastic Net model per fold of the five-fold cross-validation scheme are provided. The name of these radiomic features are provided on the left-hand side, whilst their corresponding feature classes are given on the right-hand side of the table. The AUC for each fold of cross-validation is provided and the mean AUC and standard deviation are stated in bold. The predictive performance of (A) when calcium is used as the only predictor, (B) when only radiomic features are the predictors, and (C) when using both radiomic features and calcium in an integrated model are provided. * p -value < 0.05, *** p -value < 0.001 when comparing the classification performance per fold of radiomic models B and C with the calcium only model (A) using DeLong's method. The colours highlight non-zero feature coefficients where non-zero coefficients indicate how features played a role in the model's predictions.

Using radiomic features (with resegmentation to [0, 200 HU]) alone as predictors performed better than carotid calcium alone, with a mean (SD) AUC of 0.67 (0.08) and a p -value of 0.043. The combination of radiomic features with carotid calcium as predictors led to the highest predictive performance, with a mean (SD) AUC of 0.73 (0.09), a mean (95% CI) accuracy of 69% (47–88%) and a p -value of 0.043 when compared with carotid calcium alone and a p -value of 0.042 when compared with radiomic features alone. The performance (mean AUC with SD) of the other machine learning classifiers using radiomic features and carotid calcium as predictors were: decision tree 0.58 (0.19), random forest 0.67 (0.08), LASSO 0.72 (0.09), neural network 0.60 (0.09) and XGBoost 0.56 (0.09). Please see Supplementary Table S6 for the sensitivity and specificity of each individual model.

The radiomic feature set ($n = 10$) for multi-slice analysis with resegmentation is shown in Fig. 4 along with the coefficients for each feature as determined by the Elastic Net classifier per fold within the cross-validation scheme. The feature coefficients indicate the importance of the features for the model's predictions. Larger positive coefficient values suggest higher importance for predicting the culprit carotid artery class, whilst larger negative coefficient values suggest higher importance for predicting the non-culprit carotid artery class.

Overall, there were 3 radiomic features that were highly consistent in being relevant predictors for carotid artery status across every cross-validation fold: (1) GLDM: Dependence Variance, (2) GLSZM: Grey Level NonUniformity and (3) GLRLM: Long Run High Grey Level Emphasis.

Discussion

To our knowledge, this is the first systematic approach to evaluate the robustness and reproducibility of carotid CT angiography radiomics and its impact on the ability to identify culprit carotid arteries in stroke and TIA patients. We revealed factors that impacted robustness and identified a radiomics set that could be used to predict patient symptom state. We compared our robust feature set with carotid calcium scoring using several machine learning models, demonstrating superior performance.

The ROI perturbations (morphological operations: dilations and erosions) that we used mimicked the ROI over- and under-estimation variations introduced by human subjectivity in clinical practice (inter-observer variability) when using manual delineation methods. Not all of the 93 radiomic features that we extracted were robust against these morphological perturbations and the proportion of radiomic features with excellent robustness varied depending on the image settings used.

We found that without image pre-processing, the majority of radiomic features (but not all) had excellent robustness against ROI perturbations. There were similarities between the radiomic features with poor robustness in both single-slice and multi-slice analysis to include the radiomic features: First Order: 10th Percentile and GLDM: Low Grey Level Emphasis. These radiomic features are related to low grey values within the CTA image and therefore most likely reflect the varying amounts of carotid artery perivascular fat captured in the segmentation mask following the morphological perturbations. Following grey value range resegmentation, which restricted radiomic feature calculation to Hounsfield units between 0 and 200 inclusive, the proportion of poorly robust radiomic features was greatly reduced as the low grey values that reflect perivascular fat were excluded. However, resegmentation also reduced the number of radiomic features with excellent robustness and shifted them into the moderate robustness category.

Whereas prior image normalisation appears necessary for image pre-processing of MRI scans for radiomic work (where the grey values are arbitrary), it does not seem necessary for carotid CTA scans (where grey values are calibrated to Hounsfield units). In our study, prior normalisation of CTA scans with PyRadiomics did not increase the proportion of radiomic features with excellent robustness. This is in line with most CT imaging radiomic studies that do not tend to apply prior normalisation²⁷.

The use of bin number *versus* bin width as image quantisation methods were not interchangeable. The radiomic features in one setting were not necessarily robust in another setting. We found that a fixed bin width of 25 or 30 for image quantisation led to the greatest proportion of radiomic features with excellent robustness. Since the PyRadiomics default is already a fixed bin width of 25, we recommend use of that setting in future carotid CTA radiomics studies. In the bin width settings that were investigated in this study (from 10 to 35 in increments of 5), we found a decrease in the proportion of radiomic features with excellent robustness when using the higher limit of 35, and so we did not investigate higher values than this. When using fixed bin numbers as the method for image quantisation, we found that using $BN = 256$ led to the highest proportion of poorly robust features.

For the different image settings, we investigated both a single-slice approach and a multi-slice approach. Multi-slice analysis, which involved ROI delineation of the carotid artery on multiple carotid CTA axial slices, produced a higher proportion of radiomic features with excellent robustness than single-slice analysis, which used only one CTA axial slice at the carotid artery bifurcation. Although there was a considerable overlap between the radiomic features with excellent robustness in both single-slice and multi-slice approaches, there were certain radiomic features that had excellent robustness only in either single-slice analysis or multi-slice analysis. This was also the case for radiomic features with poor robustness; certain radiomic features had excellent robustness in multi-slice analysis, but poor robustness in single-slice analysis, for example GLCM: Joint Energy. Single-slice and multi-slice analysis approaches may reveal different radiomic features that are robust because the features extracted from a single-slice are more dependent on ROI placement around the carotid bifurcation, whereas radiomic features extracted during multi-slice analysis include more voxels, capturing more information about the carotid artery. The multi-slice approach led to the highest proportion of robust radiomic features and had better predictive value than the single-slice approach. These results echo those in oncology radiomic studies, which found that using a multi-slice approach compared to a single-slice approach (i.e. whole tumour analysis *versus* the largest cross-sectional area) was more representative of tumour heterogeneity²⁸.

For the identification of culprit *versus* non-culprit carotid arteries in symptomatic patients, we investigated several machine learning algorithms which have been extensively applied to radiomics and quantitative imaging^{29–31}. This approach acknowledged the “no free lunch” theorem³²—that there is no universal best model for every task³³. In our study, the ElasticNet model achieved the highest performance amongst those investigated. We identified 10 non-redundant radiomic features with excellent robustness when using a multi-slice approach with grey value range resegmentation ($[0, 200]$ HU) that significantly outperformed carotid calcium scoring in machine learning classification. As a univariable predictor, carotid calcium had poor predictive performance. This was not surprising since there was no statistically significant difference between the carotid calcium score of culprit and non-culprit carotid arteries. Subsequently, the best predictive model consisted of the radiomic feature set with resegmentation to $[0, 200]$ HU, alongside carotid calcification. This may reflect how resegmentation excluded high grey values, largely related to carotid calcification and luminal contrast, so that differences between culprit and non-culprit carotid artery radiomic profiles could be more easily identified. In addition, this demonstrates that carotid calcium and the information captured by our radiomic features after resegmentation are complementary.

Other groups have also reported using metrics derived from CTA images, for example, Gupta et al. investigated the discriminative ability of CTA plaque thickness measurements to identify symptomatic carotid artery stenosis³⁴. This indicates that carotid CTA imaging contains information beyond luminal stenosis and in this proof-of-principle study, we have shown that radiomics is a feasible and reliable approach to extract this information.

Overall, our findings suggest that: (1) a multi-slice approach is better than a single-slice approach in terms of radiomic feature robustness and predictive accuracy, (2) there is no need for image normalisation in carotid CTA radiomic studies, (3) grey value range resegmentation can help improve predictive accuracy and (4) because radiomic features can be susceptible to changes in the imaging and radiomics workflow, it is important that future studies include detailed descriptions of the image settings used to ensure reproducibility and replicability. This information would ideally include the image acquisition protocols, image pre-processing details and method of interpolation, method and value of image quantisation, and radiomic feature definitions.

Limitations

One limitation of this study is its retrospective nature—the carotid imaging datasets were pooled from three prior vascular imaging studies: ICARUSS³⁵, VISION³⁶ and CHAI³⁷. Additionally, all images were acquired using the same scanner in one centre. Consequently, the robust radiomic features identified here may be specific to datasets derived in similar settings. In addition, our imaging dataset captured information from culprit carotid arteries after plaque rupture had occurred. Ideally, we would highlight high risk arteries before that stage. Now that we have identified CT-based carotid radiomic features that are robust, a prospective study of at-risk patients using different hardware manufacturers will be an important next step.

We also acknowledge that the use of a 3 mm slice thickness may result in loss of some information and might lead to partial volume effects. Nevertheless, even using this slice thickness, we did identify a robust subset of features that could classify carotid plaques with reasonable accuracy. Further work should test the hypothesis that thinner image slices perform better.

Another consideration is that we used anatomical criteria to standardise the region of the carotid artery that was segmented for single-slice analysis (i.e. the axial slice through the bifurcation) and multi-slice analysis (14 slices about the carotid bifurcation). A possible limitation of this approach, particularly for single-slice analysis, is that the responsible carotid plaque may not be fully captured in the ROI. To account for this, we tested the predictive performance of a multi-slice analysis. We also investigated the impact of prior resegmentation of the image to limit the HU values analysed and counter possible differences in arterial contrast densities.

Finally, here we investigated only unfiltered radiomic features. There are other radiomic parameters that can be extracted after image filtering, such as Gabor filters and wavelet transformations. However, as this was a first proof-of-principle study using first-order and higher-order radiomic features, we wanted to limit the number of features extracted. Future work could expand on this. As the primary objective of this study was robustness analysis rather than developing a definitive radiomics signature, the default Python scikit-learn configurations for the machine learning classifiers were used, without extensive hyperparameter tuning. This avoided further reduction of the limited dataset that could be used for training the machine learning classifiers. Nevertheless, it may be that the predictive performance we have already achieved could be bettered with hyperparameter tuning in future work.

Conclusion

In summary, to the best of our knowledge, this is the first systematic approach to evaluate the robustness and reproducibility of CT radiomics in carotid artery atherosclerosis. We identified a set of radiomic features that are robust, non-redundant and have superior predictive performance, over and above the degree of calcification, for the classification of culprit *versus* non-culprit carotid arteries in patients with stroke and TIA. If validated prospectively, this carotid CT radiomic features set could improve stroke prediction and target therapies to those at highest risk.

Methods

Carotid CT dataset. This study used carotid CTA scans pooled from three observational vascular imaging research datasets from a single institution (Addenbrooke's Hospital, Cambridge University Hospitals National Health Service Foundation Trust, Cambridge, UK)^{35–37}. All studies had appropriate ethical approvals in place by the Cambridge Central Research Ethics Committee; informed consent was obtained from all patients and the studies were conducted according to relevant guidelines and regulations. The studies had similar inclusion and exclusion criteria, which are listed in the published papers^{35–37}. All participants had experienced a carotid artery-related ischaemic stroke or TIA during the 3 months before imaging.

In total, data from 41 patients were included, comprising 82 carotid arteries (41 culprit and 41 non-culprit). The culprit carotid artery was determined by the side consistent with the clinical presentation of stroke (or TIA) symptoms, and the non-culprit carotid artery was defined as the artery contralateral to the culprit. Further details of how the culprit carotid plaque was identified and how carotid images with and without contrast were acquired using a standard clinical protocol are described in Supplementary Methods S1.

Image analysis. Figure 5 illustrates the radiomics workflow within this study. All CT images were analysed by a reader (EPVL) blinded to the clinical status of the carotid artery. Where a second reader is mentioned, (CW), they were also blinded in the same fashion. Details of the methodology used to assess carotid artery plaque characteristics are found in Supplementary Methods S2, and details of intra- and inter-observer reproducibility evaluation are provided in Supplementary Methods S3.

Manual segmentation: single-slice analysis. In single-slice analysis, one axial CTA slice, at the carotid bifurcation, was used on each side, with original slice thickness of 0.625 mm and slice spacing of 0.4 mm. ROIs were drawn to encompass the whole vessel as closely as possible, including the outer wall, using commercially available research software (TexRad; Feedback Medical Ltd, Cambridge, UK).

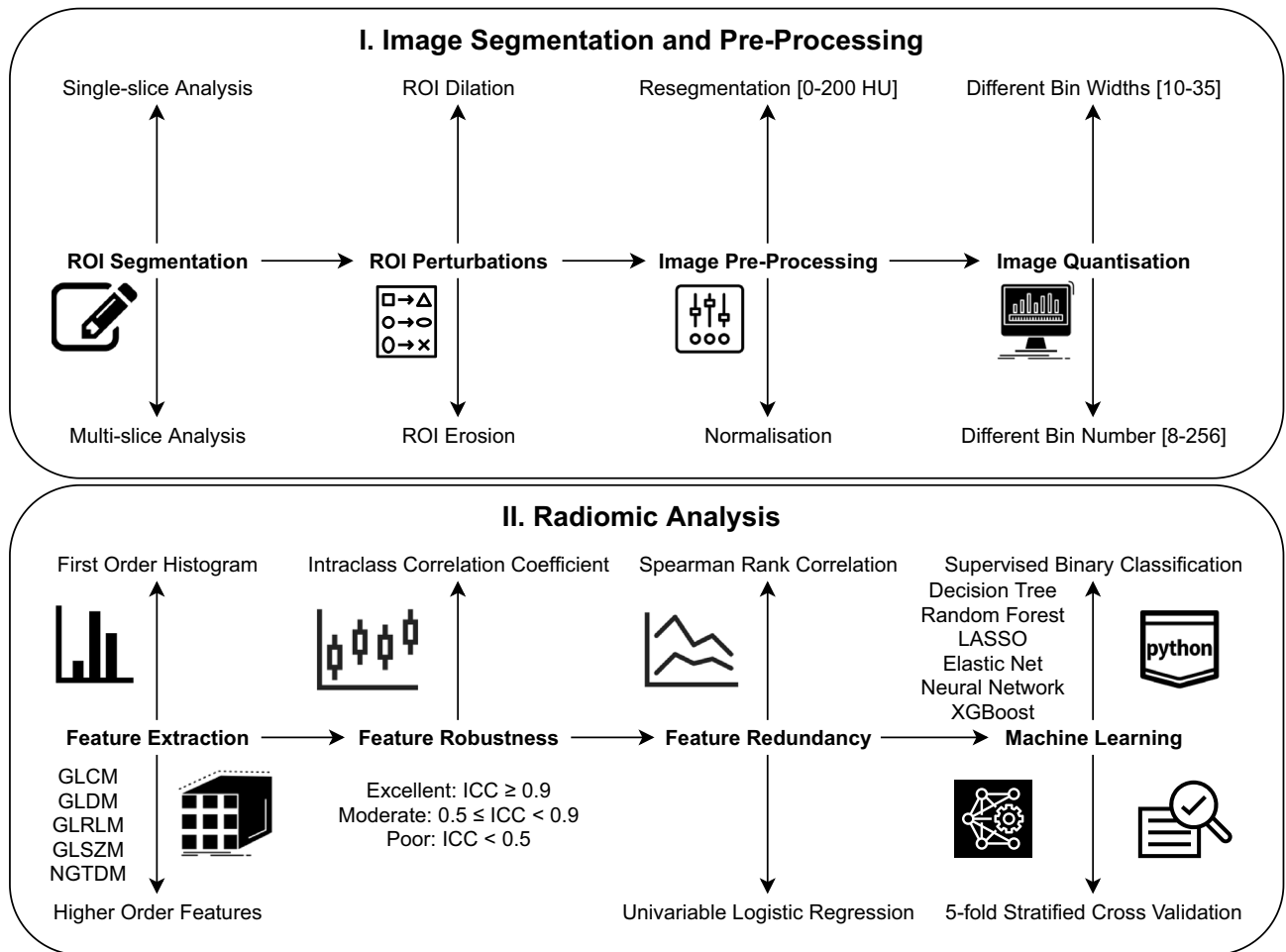


Figure 5. Radiomics workflow. The upper panel illustrates the steps taken from manual segmentation of the carotid CTA images to create ROIs for single-slice analysis (and VOIs for multi-slice analysis) to segmentation mask perturbations, prior image normalisation or resegmentation and image quantisation. The lower panel outlines the subsequent process of radiomic features extraction, robustness analysis and machine learning for the differentiation of culprit versus non-culprit carotid arteries. The predictive ability of the classifiers was assessed via five-fold stratified cross-validation.

Manual segmentation: multi-slice analysis. CTA slices were resampled to 3 mm slice thickness using the OsiriX MD software resampling plugin (Pixmeo SARL, Bernex, Geneva, Switzerland) as per published methods^{35–37}. 14 consecutive carotid artery slices were manually segmented using TexRad (as in single-slice analysis) with ROIs drawn around the carotid artery adventitia, with the carotid bifurcation designated as slice zero^{35–37}. Reads incorporated all slices from 3 below the carotid bifurcation to 10 slices above, covering portions of the common carotid and internal carotid arteries. For each carotid artery, the 14 consecutive slices were amalgamated into a single VOI from which radiomic features were subsequently extracted.

Radiomics feature extraction. PyRadiomics is an open-source Python package developed for the standardisation of radiomic feature extraction³⁸. PyRadiomics and Python were used for feature extraction from the ROIs and VOIs described above. Six feature classes were extracted: (1) first-order intensity histogram statistics, (2) Grey Level Co-occurrence Matrix features (GLCM)^{39,40}, (3) Grey Level Run Length Matrix features (GLRLM)⁴¹, (4) Grey Level Size Zone Matrix (GLSZM)⁴², (5) Grey Level Dependence Matrix (GLDM)⁴³ and (6) Neighbouring Grey Tone Difference Matrix Features (NGTDM)⁴⁴. Please see Supplementary Table S7 and S8 for details of the individual extracted radiomic features.

Robustness analysis. ROI perturbations. Manual segmentation (as opposed to automatic segmentation) is a source of intra- and inter-observer variability. Automatic segmentation methods are not currently widely available in medicine, although this is an area of active development. We therefore evaluated the impact of perturbations to ROI delineation on the extracted radiomic features by systematically performing ROI dilation and erosion. This was to simulate certain variations in ROI/VOI placement that may occur in clinical practice, including over-estimation (with dilation), and under-estimation (with erosion).

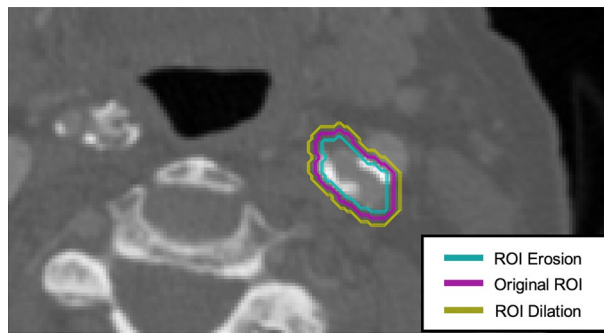


Figure 6. ROI segmentation and perturbations. Carotid CTA images were manually segmented to delineate the carotid artery. The original ROI was subjected to morphological operations: erosions and dilations aimed to assess the robustness of radiomic features to perturbations in image segmentation.

To achieve these perturbations, the original ROIs delineated by the primary reader (EPVL) were subjected to the dilation and erosion image morphological operations implemented in Python, see Fig. 6.

For single-slice analysis, we used a circular structuring element of radius 1, with iterations of 1–2 for ROI dilation and erosion. For multi-slice analysis, we used a spherical structuring element of radius 1, with iterations of 1–2 for ROI dilation, but only 1 iteration for ROI erosion in order to ensure that a sufficient number of pixels would be available for the downstream radiomic feature calculation after erosion. Where resegmentation was applied as a pre-processing scheme, ROI erosion was not performed, only ROI dilation to ensure that all ROIs had sufficient pixels for radiomic feature extraction, details of resegmentation are provided below.

Image pre-processing. Prior to radiomic feature calculations, there are different image pre-processing schemes that can be applied to a CTA scan. Three schemes were investigated: (a) Original image (no image pre-processing applied), (b) Normalisation and (c) Resegmentation.

Normalisation is generally a necessary image pre-processing step for magnetic resonance images since their grey values are arbitrary. In contrast, the grey values in CT images are already calibrated to HUs. However, CTA images may have differences in contrast filling and so we investigated the impact of prior image normalisation to the robustness of the extracted radiomic features. When investigating the image normalisation scheme, the CTA image was normalised such that the pixel values assumed an approximate Gaussian distribution.

Resegmentation refers to the process whereby only pixels within a specified grey value range are retained for radiomic feature calculation within the ROI/VOI⁴⁵. Resegmentation was applied with an upper limit of 200, and a lower limit of 0 which restricted radiomic feature extraction to only the pixels with HU values between 0 and 200. This grey value range resegmentation aids with excluding the effects of excess carotid macro-calcification and limits the effect of luminal contrast and perivascular carotid fat within the CTA ROI/VOI. For resegmentation, we used a fixed BW of 25 only (PyRadiomics version 3.0 default) for image quantisation, described further below.

Image quantisation. Image quantisation refers to the conversion of image grey values to a discrete set of grey value counts. Before radiomic features are calculated, the image must be quantised by using a fixed number of bins, or by using a fixed BW. We varied the BWs of the image grey value histogram from 10 to 35, in increments of 5. For BN variations, we varied the fixed number of bins as follows: 8, 16, 32, 64, 128 and 256. This range of bin sizes was chosen based on the guidance in the PyRadiomics documentation^{46,47}.

Multi-slice analysis: Image resampling and interpolation method. Higher-order radiomic feature extraction requires isotropic images, i.e. the pixel dimensions in the x , y and z directions are the same, to be rotationally invariant^{48,49}. In CT imaging, images are often isotropic in-plane but will have a larger z -axis slice spacing and therefore be anisotropic in 3D. In radiomics studies, it is common for images to be isotropically resampled. We investigated the effect of using B-spline interpolation (PyRadiomics default) *versus* linear interpolation (faster and simpler) to resample the 3 mm slice thickness images and VOI segmentation masks to $1 \times 1 \times 1 \text{mm}^3$ on the extracted radiomic features.

Statistical analysis. For statistical comparisons between culprit *versus* non-culprit carotid arteries, the difference between the two paired groups were assessed for normality visually with histogram plots and statistically with the Shapiro–Wilk test. Where the normality assumption was met, the paired t -test was used, if not, the non-parametric Wilcoxon signed-rank test was used. A p -value < 0.05 was considered statistically significant.

The Dice coefficient (DC), a measure of segmentation overlap commonly used in computer vision and machine learning applications⁵⁰, was calculated to assess agreement between ROI segmentations in the following ways: (1) comparing the ROIs for 8 carotid arteries drawn by the primary reader (EPVL) at two separate time points to determine intra-observer variability, (2) comparing ROIs for 8 carotid arteries drawn by the primary reader with those drawn by a second independent reader (CW) to determine inter-observer variability and (3) comparing the ROIs for 82 carotid arteries drawn by the primary reader with the ROIs generated following

morphological operations (dilations and erosions) to determine the variability generated by systematic ROI perturbations.

The DC measures the level of agreement between different image segmentations by considering the level of overlap between ROI X and ROI Y over the total number of pixels in ROI X and ROI Y according to Eq. (1):

$$DC = \frac{2|X \cap Y|}{|X| + |Y|}, \quad (1)$$

where $|\cdot|$ denotes the cardinality of the pixels contained in a certain set.

We measured the degree of robustness using the 2-way mixed-effects model, absolute agreement, single rater and the 2-way mixed-effects model, consistency, single rater intraclass correlation coefficient (ICC) according to the McGraw and Wong convention⁵¹ and according to the ICC guidelines of Koo and Li⁵², as appropriate.

Let n and k be the number of subjects and number of raters/measurements, respectively, the ICCs used are defined as follows:

two-way mixed effects, consistency, single rater/measurement:

$$ICC(3, 1) = \frac{MS_R - MS_E}{MS_R + (k - 1)MS_E}; \quad (2)$$

two-way mixed effects, absolute agreement, single rater/measurement:

$$ICC(2, 1) = \frac{MS_R - MS_E}{MS_R + (k - 1)MS_E + \frac{k}{n}(MS_C - MS_E)}; \quad (3)$$

where MS_R , MS_E and MS_C are the mean square for rows, mean square for error and mean square for columns, respectively.

The ICC values fall between 0 and 1. Radiomic features were classified into three groups, with ICC values < 0.5 , between 0.5 to 0.9, and ≥ 0.9 , being indicative of poor, moderate and excellent robustness, respectively⁵³.

All statistical analysis was performed in IBM SPSS Statistics for Macintosh and Python. Further details about the software and packages used are provided in Supplementary Methods S4.

Machine learning classification. Only the features with excellent robustness were used for the classification of culprit *versus* non-culprit carotid arteries. To reduce multicollinearity and feature redundancy, pairwise feature-to-feature correlations were determined using the Spearman Rank correlation. For pairs of features with a $Spearman|r_s| \geq 0.95$, the feature with the highest AUC in univariate logistic regression was retained, and the latter was discarded⁵⁴.

The features were subsequently standardised to have a mean of zero and a variance of one. 6 machine learning classifiers were evaluated, using a random state of 42 for reproducibility: decision tree⁵⁵, random forest⁵⁶, LASSO regression⁵⁷, Elastic Net regression (weight for L_1 and L_2 penalties = 0.5)⁵⁸, a neural network⁵⁹ and XGBoost⁶⁰. Further details about the machine learning classifier configurations are provided in Supplementary Methods S5. The dataset was shuffled and the average performance (accuracy and AUC) of the classifiers calculated following five-fold stratified cross-validation.

The AUC of the radiomics-only models, and of the integrated models (using radiomics features and calcium as predictors) were compared with the AUC of the calcium-only models using DeLong's method⁶¹ to compare classifier performance for both single- and multi-slice approaches in each fold of the five-fold cross-validation scheme. The distribution of AUC values was compared using the Wilcoxon signed-rank test for the following comparisons: (1) calcium-only *versus* radiomics-only model, (2) calcium-only *versus* integrated model and (3) radiomics-only *versus* integrated model.

Data availability

The anonymised datasets used in the current study are available from the corresponding author upon reasonable request.

Received: 14 October 2020; Accepted: 21 January 2021

Published online: 10 February 2021

References

- Karlsson, L. *et al.* Risk of recurrent stroke in patients with symptomatic mild (20–49% NASCET) carotid artery stenosis. *Eur. J. Vasc. Endovasc. Surg.* **52**, 287–294 (2016).
- Gillies, R. J., Kinahan, P. E. & Hricak, H. Radiomics: images are more than pictures, they are data. *Radiology* **278**, 563–577 (2016).
- Rizzo, S. *et al.* CT radiogenomic characterization of EGFR, K-RAS, and ALK mutations in non-small cell lung cancer. *Eur. Radiol.* **26**, 32–42 (2016).
- Zhang, Y., Oikonomou, A., Wong, A., Haider, M. A. & Khalvati, F. Radiomics-based prognosis analysis for non-small cell lung cancer. *Sci. Rep.* **7**, 46349 (2017).
- Aerts, H. J. W. L. *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* **5**, 4006 (2014).
- Lambin, P. *et al.* Radiomics: the bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* **14**, 749–762 (2017).

7. AJCC Cancer Staging Atlas. Springer. New York. <https://doi.org/10.1007/978-1-4614-2080-4> (2012).
8. Mannil, M. *et al.* Texture analysis of myocardial infarction in CT: comparison with visual analysis and impact of iterative reconstruction. *Eur. J. Radiol.* **113**, 245–250 (2019).
9. Nam, K. *et al.* Value of computed tomography radiomic features for differentiation of periprosthetic mass in patients with suspected prosthetic valve obstruction. *Circ. Cardiovasc. Imaging* **12**, (2019).
10. Neisius, U. *et al.* Radiomic analysis of myocardial Native T1 imaging discriminates between hypertensive heart disease and hypertrophic cardiomyopathy. *JACC Cardiovasc. Imaging* **12**, 1946–1954 (2019).
11. Kolossváry, M. *et al.* Radiomic features are superior to conventional quantitative computed tomographic metrics to identify coronary plaques with napkin-ring sign. *Circ. Cardiovasc. Imaging* **10**, (2017).
12. Kyriacou, E. *et al.* Classification of atherosclerotic carotid plaques using morphological analysis on ultrasound images. *Appl. Intell.* **30**, 3–23 (2009).
13. Awad, J., Krasinski, A., Parraga, G. & Fenster, A. Texture analysis of carotid artery atherosclerosis from three-dimensional ultrasound images: Texture analysis of carotid artery atherosclerosis. *Med. Phys.* **37**, 1382–1391 (2010).
14. Prescott, J. W. Quantitative imaging biomarkers: the application of advanced image processing and analysis to clinical and preclinical decision making. *J. Digit. Imaging* **26**, 97–108 (2013).
15. Berenguer, R. *et al.* Radiomics of CT features may be nonreproducible and redundant: influence of CT acquisition parameters. *Radiology* **288**, 407–415 (2018).
16. Mackin, D. *et al.* Measuring computed tomography scanner variability of radiomics features. *Invest. Radiol.* **50**, 757–765 (2015).
17. Zwanenburg, A. *et al.* The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* **295**, 328–338 (2020).
18. Zwanenburg, A. *et al.* Assessing robustness of radiomic features by image perturbation. *Sci. Rep.* **9**, 614 (2019).
19. Shafiq-ul-Hassan, M. *et al.* Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Med. Phys.* **44**, 1050–1062 (2017).
20. Buch, K., Kuno, H., Qureshi, M. M., Li, B. & Sakai, O. Quantitative variations in texture analysis features dependent on MRI scanning parameters: a phantom model. *J. Appl. Clin. Med. Phys.* **19**, 253–264 (2018).
21. Traverso, A., Wee, L., Dekker, A. & Gillies, R. Repeatability and reproducibility of radiomic features: a systematic review. *Int. J. Radiat. Oncol.* **102**, 1143–1158 (2018).
22. Zhao, B. *et al.* Reproducibility of radiomics for deciphering tumor phenotype with imaging. *Sci. Rep.* **6**, 23428 (2016).
23. Schwier, M. *et al.* Repeatability of multiparametric prostate MRI radiomics features. *Sci. Rep.* **9**, 9441 (2019).
24. Hatt, M., Tixier, F., Cheze Le Rest, C., Pradier, O. & Visvikis, D. Robustness of intratumour 18F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma. *Eur. J. Nucl. Med. Mol. Imaging* **40**, 1662–1671 (2013).
25. Edalat-Javid, M. *et al.* Cardiac SPECT radiomic features repeatability and reproducibility: a multi-scanner phantom study. *J. Nucl. Cardiol.* <https://doi.org/10.1007/s12350-020-02109-0> (2020).
26. Jang, J. *et al.* Reproducibility of myocardial radiomic features in cardiac MRI imaging. *Radiol. Cardiothorac. Imaging* (2020).
27. Larue, R. T. H. M., Defraene, G., De Ruysscher, D., Lambin, P. & van Elmpt, W. Quantitative radiomics studies for tissue characterization: a review of technology and methodological procedures. *Br. J. Radiol.* **90**, 20160665 (2017).
28. Ng, F., Kozarski, R., Ganeshan, B. & Goh, V. Assessment of tumor heterogeneity by CT texture analysis: Can the largest cross-sectional area be used as an alternative to whole tumor analysis?. *Eur. J. Radiol.* **82**, 342–348 (2013).
29. Parmar, C. *et al.* Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer. *Front. Oncol.* **5**, (2015).
30. Hajianfar, G. *et al.* Noninvasive O6 methylguanine-DNA methyltransferase status prediction in glioblastoma multiforme cancer using magnetic resonance imaging radiomics features: univariate and multivariate radiogenomics analysis. *World Neurosurg.* **132**, e140–e161 (2019).
31. Leger, S. *et al.* A comparative study of machine learning methods for time-to-event survival data for radiomics risk modelling. *Sci. Rep.* **7**, 13206 (2017).
32. Ho, Y. C. & Pepyne, D. L. Simple explanation of the no-free-lunch theorem and its implications. *J. Optim. Theory Appl.* **115**, 549–570 (2002).
33. Shiri, I. *et al.* Next-generation radiogenomics sequencing for prediction of EGFR and KRAS mutation status in NSCLC patients using multimodal imaging and machine learning algorithms. *Mol. Imaging Biol.* **22**, 1132–1148 (2020).
34. Gupta, A. *et al.* Evaluation of computed tomography angiography plaque thickness measurements in high-grade carotid artery stenosis. *Stroke* **45**, 740–745 (2014).
35. Evans, N. R. *et al.* Dual-tracer positron-emission tomography for identification of culprit carotid plaques and pathophysiology in vivo. *Circ. Cardiovasc. Imaging* **13**, (2020).
36. Tarkin, J. M. *et al.* Detection of atherosclerotic inflammation by 68 Ga-DOTATATE PET compared to [18 F]FDG PET imaging. *J. Am. Coll. Cardiol.* **69**, 1774–1791 (2017).
37. Joshi, F. R. *et al.* Vascular imaging with 18 F-fluorodeoxyglucose positron emission tomography is influenced by hypoxia. *J. Am. Coll. Cardiol.* **69**, 1873–1874 (2017).
38. van Griethuysen, J. J. M. *et al.* Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* **77**, e104–e107 (2017).
39. Haralick, R. M., Shanmugam, K. & Dinstein, I. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* **SMC-3**, 610–621 (1973).
40. Haralick, R. M. Statistical and structural approaches to texture. *Proc. IEEE* **67**, 786–804 (1979).
41. Galloway, M. M. Texture analysis using gray level run lengths. *Comput. Graph. Image Process.* **4**, 172–179 (1975).
42. Thibault, G., Angulo, J. & Meyer, F. Advanced statistical matrices for texture characterization: application to cell classification. *IEEE Trans. Biomed. Eng.* **61**, 630–637 (2014).
43. Sun, C. & Wee, W. G. Neighboring gray level dependence matrix for texture classification. *Comput. Vis. Graph. Image Process.* **23**, 341–352 (1983).
44. Amadasun, M. & King, R. Textural features corresponding to textural properties. *IEEE Trans. Syst. Man Cybern.* **19**, 1264–1274 (1989).
45. Ibrahim, A. *et al.* Radiomics analysis for clinical decision support in nuclear medicine. *Semin. Nucl. Med.* **49**, 438–449 (2019).
46. Frequently Asked Questions—pyradiomics v3.0.post5+gf06ac1d documentation. <https://pyradiomics.readthedocs.io/en/latest/faq.html?highlight=bin%20width#what-about-gray-value-discretization-fixed-bin-width-fixed-bin-count>.
47. Tixier, F. *et al.* Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. *J. Nucl. Med.* **52**, 369–378 (2011).
48. Forghani, R. *et al.* Radiomics and artificial intelligence for biomarker and prediction model development in oncology. *Comput. Struct. Biotechnol. J.* **17**, 995–1008 (2019).
49. Feng, B. *et al.* Radiomics nomogram for preoperative differentiation of lung tuberculoma from adenocarcinoma in solitary pulmonary solid nodule. *Eur. J. Radiol.* **128**, 109022 (2020).
50. Zou, K. H. *et al.* Statistical validation of image segmentation quality based on a spatial overlap index1. *Acad. Radiol.* **11**, 178–189 (2004).
51. McGraw, K. O. & Wong, S. P. Forming inferences about some intraclass correlation coefficients. *Psychol. Methods* **1**, 30–46 (1996).

52. Koo, T. K. & Li, M. Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **15**, 155–163 (2016).
53. Cattell, R., Chen, S. & Huang, C. Robustness of radiomic features in magnetic resonance imaging: review and a phantom study. *Vis. Comput. Ind. Biomed. Art* **2**, 19 (2019).
54. Owens, C. A. *et al.* Lung tumor segmentation methods: Impact on the uncertainty of radiomics features for non-small cell lung cancer. *PLoS ONE* **13**, e0205003 (2018).
55. Gordon, A. D., Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J. Classification and Regression Trees. *Biometrics* **40**, 874 (1984).
56. Tin Kam Ho. Random decision forests. in *Proceedings of 3rd International Conference on Document Analysis and Recognition* vol. 1 278–282 (IEEE Comput. Soc. Press, 1995).
57. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Methodol.* **58**, 267–288 (1996).
58. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67**, 301–320 (2005).
59. Bishop, C. M. *Neural networks for pattern recognition*. (Clarendon Press ; Oxford University Press, 1995).
60. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (ACM, 2016). <https://doi.org/10.1145/2939672.2939785>.
61. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44**, 837–845 (1988).

Acknowledgements

EPVL is undertaking a PhD funded by the Cambridge School of Clinical Medicine, Frank Edward Elmore Fund and the Medical Research Council's Doctoral Training Partnership [award reference: 1966157]. JMT is supported by a Wellcome Trust Clinical Research Career Development Fellowship [211100/Z/18/Z], the National Institute for Health Research (NIHR) Imperial Biomedical Research Centre and the British Heart Foundation Cambridge Centre of Research Excellence. NRE was supported by a Research Training Fellowship from The Dunhill Medical Trust [RTF44/0114]. MMC was supported by fellowships from the Royal College of Surgeons of England, and the British Heart Foundation [BHF; FS/16/29/31957]. HP is undertaking a PhD with a BHF CRE studentship. FJG is an NIHR Senior Investigator. LR and ES were supported by The Mark Foundation for Cancer Research and Cancer Research UK (CRUK) Cambridge Centre [C9685/A25177]. MR is supported by AstraZeneca Oncology R&D. ES receives additional support provided by the NIHR Cambridge Biomedical Research Centre. FAG receives funding from CRUK. EAW receives support from the NIHR CRN. CBS acknowledges support from the Leverhulme Trust project on 'Breaking the non-convexity barrier', the Philip Leverhulme Prize, the EPSRC grants EP/S026045/1 and EP/T003553/1, the EPSRC Centre Nr. EP/N014588/1, the Wellcome Innovator Award RG98755, European Union Horizon 2020 research and innovation programmes under the Marie Skłodowska-Curie grant agreement No. 777826 NoMADS and No. 691070 CHiPS, the Cantab Capital Institute for the Mathematics of Information and the Alan Turing Institute. JHFR is part-supported by the NIHR Cambridge Biomedical Research Centre, the British Heart Foundation, HEFCE, the Wellcome Trust and the EPSRC grant [EP/N014588/1] for the University of Cambridge Centre for Mathematical Imaging in Healthcare. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Author contributions

E.P.V.L., L.R., F.Z., Y.H., J.H.F.R. were involved with the conception of the work. E.P.V.L., L.R., A.L., F.A.G., C.B.S., E.S. and J.H.F.R. were involved with the design of the work. J.M.T., N.R.E., M.M.C., P.A.C., E.A.W. and J.H.F.R. were involved with the acquisition of the data. E.P.V.L., C.W., R.S. were involved with the analysis of the data. E.P.V.L., L.R., H.P., J.R.W., J.H.F.R. were involved with the interpretation of data. E.P.V.L. and J.H.F.R. have drafted the manuscript. L.R., A.L., H.P., M.R., F.J.G. and F.A.G. have substantially revised the work. All authors reviewed the work, contributed to its overall scientific content and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-82760-w>.

Correspondence and requests for materials should be addressed to J.H.F.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021