European Journal of Radiology Open 13 (2024) 100594

Contents lists available at ScienceDirect

European Journal of Radiology Open

journal homepage: www.elsevier.com/locate/ejro

Using machine learning to predict carotid artery symptoms from CT angiography: A radiomics and deep learning approach

Elizabeth P.V. Le [a,1], Mark Y.Z. Wong [a,*,1], Leonardo Rundo [b,c,d], Jason M. Tarkin [a], Nicholas R. Evans [e], Jonathan R. Weir-McCall [b,f], Mohammed M. Chowdhury [g], Patrick A. Coughlin [m], Holly Pavey [h], Fulvio Zaccagna [b,i,j], Chris Wall [a], Rouchelle Sriranjan [a], Andrej Corovic [a], Yuan Huang [a,b,k], Elizabeth A. Warburton [e], Evis Sala [n,o], Michael Roberts [a,k,l], Carola-Bibiane Schönlieb [k,1], James H.F. Rudd [a,k]

[a] Department of Medicine, University of Cambridge, United Kingdom
[b] Department of Radiology, University of Cambridge, United Kingdom
[c] Cancer Research UK Cambridge Centre, University of Cambridge, United Kingdom
[d] Department of Information and Electrical Engineering and Applied Mathematics (DIEM), University of Salerno, Italy
[e] Department of Clinical Neurosciences, University of Cambridge, United Kingdom
[f] Department of Radiology, Royal Papworth Hospital, Cambridge, UK
[g] Division of Vascular Surgery, Department of Surgery, University of Cambridge, United Kingdom
[h] Division of Experimental Medicine and Immunotherapeutics, University of Cambridge, United Kingdom
[i] Department of Imaging, Cambridge University Hospitals NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge, United Kingdom
[j] Investigative Medicine Division, Radcliffe Department of Medicine, University of Oxford, Oxford, United Kingdom
[k] EPSRC Centre for Mathematical Imaging in Healthcare, University of Cambridge, United Kingdom
[l] Department of Applied Mathematics and Theoretical Physics, University of Cambridge, United Kingdom
[m] Department of Vascular Surgery, University of Leeds, United Kingdom
[n] Dipartimento di Scienze Radiologiche ed Ematologiche, Università Cattolica del Sacro Cuore, Rome, Italy
[o] Dipartimento Diagnostica per Immagini, Radioterapia Oncologica ed Ematologia, Policlinico Universitario A. Gemelli IRCCS, Rome, Italy

HIGHLIGHTS

- We apply novel image analysis techniques in radiomics and deep learning.
- These techniques can identify carotid artery disease from CT images.
- Radiomics and deep learning approaches outperformed the calcium score.
- These techniques may facilitate better stroke risk classification.

ARTICLE INFO

Keywords:
AI
Stroke
Radiomics
Machine learning
Carotid artery
Coronary calcium

ABSTRACT

Purpose: To assess radiomics and deep learning (DL) methods in identifying symptomatic Carotid Artery Disease (CAD) from carotid CT angiography (CTA) images. We further compare the performance of these novel methods to the conventional calcium score.

Methods: Carotid CT angiography (CTA) images from symptomatic patients (ischaemic stroke/transient ischaemic attack within the last 3 months) and asymptomatic patients were analysed. Carotid arteries were classified into culprit, non-culprit and asymptomatic. The calcium score was assessed using the Agatston method. 93 radiomic features were extracted from regions-of-interest drawn on 14 consecutive CTA slices. For DL,

* Corresponding author.
E-mail addresses: epvl2@cam.ac.uk (E.P.V. Le), markwongyz@gmail.com (M.Y.Z. Wong), lrundo@unisa.it (L. Rundo), jt545@cam.ac.uk (J.M. Tarkin), ne214@cam.ac.uk (N.R. Evans), jw2079@cam.ac.uk (J.R. Weir-McCall), mmc59@cam.ac.uk (M.M. Chowdhury), pnk22@cam.ac.uk (P.A. Coughlin), hp409@medschl.cam.ac.uk (H. Pavey), fz247@cam.ac.uk (F. Zaccagna), cw463@cam.ac.uk (C. Wall), rfss2@cam.ac.uk (R. Sriranjan), ac476@cam.ac.uk (A. Corovic), yh288@cam.ac.uk (Y. Huang), eaw23@medschl.cam.ac.uk (E.A. Warburton), es220@medschl.cam.ac.uk (E. Sala), mr808@cam.ac.uk (M. Roberts), cbs31@cam.ac.uk (C.-B. Schönlieb), jhfr2@cam.ac.uk (J.H.F. Rudd).
[1] The above authors contributed equally.

https://doi.org/10.1016/j.ejro.2024.100594
Received 24 April 2024; Received in revised form 20 July 2024; Accepted 4 August 2024
2352-0477/© 2024 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

convolutional neural networks (CNNs) with and without transfer learning were trained directly on CTA slices. Predictive performance was assessed over 5-fold cross validated AUC scores. SHAP and GRAD-CAM algorithms were used for explainability.

Results: 132 carotid arteries were analysed (41 culprit, 41 non-culprit, and 50 asymptomatic). For asymptomatic vs symptomatic arteries, radiomics attained a mean AUC of 0.96(± 0.02), followed by DL 0.86(± 0.06) and then calcium 0.79(± 0.08). For culprit vs non-culprit arteries, radiomics achieved a mean AUC of 0.75(± 0.09), followed by DL 0.67(± 0.10) and then calcium 0.60(± 0.02). For multi-class classification, the mean AUCs were 0.95(± 0.07), 0.79(± 0.05), and 0.71(± 0.07) for radiomics, DL and calcium, respectively. Explainability revealed consistent patterns in the most important radiomic features.

Conclusions: Our study highlights the potential of novel image analysis techniques in extracting quantitative information beyond calcification in the identification of CAD. Though further work is required, the transition of these novel techniques into clinical practice may eventually facilitate better stroke risk stratification.

1. Introduction

Carotid Artery Disease (CAD) plays a major role in the development of ischaemic stroke [1]. 1 in 4 stroke survivors will have another stroke within five years, with most recurrences within one year [2]. This represents a critical window to reduce future stroke risk. Carotid endarterectomy is a surgical procedure that reduces the risk of stroke in participants with significant stenosis and is currently indicated based on the degree of carotid stenosis and the presence of relevant symptoms [3,4].

Whilst a known marker of atherosclerotic disease, the degree of stenosis alone fails to adequately identify 'vulnerable' patients. Stenosis does not inform about underlying plaque remodelling, plaque instability, or the degree of intimal inflammation; all processes which may spare luminal patency [5,6]. Furthermore, whilst the criteria for carotid revascularisation focuses on 70 % carotid stenosis ('high-grade carotid stenosis'), the prevalence of such a degree of stenosis is low, and a high burden of disease may still be present in the absence of high-grade stenosis [7]. Similarly, while carotid artery calcification is another common marker of atherosclerosis, the association between carotid calcification and stroke risk is likewise unclear, and the quantification methods are not as standardised as their coronary counterparts [8,9]. Better tools are needed to characterise stroke risk.

Novel image analysis methods, such as radiomics and deep learning (DL), enable the extraction of higher-dimensional data from CTA scans that may be predictive of future vascular events [10]. Radiomics is a feature-based approach that extracts quantitative image-derived metrics that can then be fed into diagnostic and prognostic prediction models [10,11]. Similarly, convolutional neural networks (CNNs) are specialised DL architectures that extract key spatial features from input images, mimicking the functionality of biological photoreceptor receptive fields [12]. In contrast to radiomics, CNNs do not require user-defined features for extraction, and are capable of automatically identifying de-novo features relevant to the specified task - such as the identification of vulnerable carotid arteries [13,14]. Both techniques enable the extraction of information from CTA scans beyond carotid luminal stenosis or calcification.

This is a "proof of concept" study that aims to evaluate the utility of radiomics and DL methods in the risk stratification of CAD. We use a machine learning approach to assess the discriminative ability of these image analysis techniques in identifying symptomatic CAD from carotid CTA images. We then compare the performance of these methods to conventional methods of atherosclerotic assessment such as the calcium score.

2. Methods

2.1. Data extraction and collection

2.1.1. Study population and data

This retrospective study analysed unenhanced CT and CTA images pooled from three previous observational research vascular imaging datasets (ICARUSS, VISION and CHAI) originating from a single institution (Addenbrooke's Hospital, Cambridge, UK) between 2011 and 2016. Details on the methodology, ethical approvals and inclusion and exclusion criteria have been published previously [15,16].

Imaging was performed on a combined GE Discovery 690 PET-CT scanner with an integrated 64-slice CT scanner (GE Healthcare, Waukesha, WI, USA), although PET imaging data was not included in this post-hoc analysis of CT images. CTA images were acquired from the aortic arch to the circle of Willis, using bolus tracking with a region-of-interest placed in the aortic arch. CTA acquisition parameters have been detailed previously [15,16].

2.2. Classification of arteries

Symptomatic (Sx) patients were defined as patients with confirmed carotid artery-related ischaemic stroke or TIA within 3 months before imaging. The carotid artery associated with precipitating the cerebrovascular event was deemed the 'culprit' (CC) carotid artery and determined via consistency with the clinical presentation of stroke (or TIA) symptoms, whilst the contralateral carotid artery was deemed the 'non-culprit' (NC) carotid artery. Carotid arteries with prior endarterectomy were excluded from analysis. Asymptomatic (Asx) patients in this study were patients with no prior TIA or stroke. Details on culprit carotid plaque identification have also been previously described [15,16].

2.3. Carotid artery analysis

Carotid arteries were evaluated both as single-slice and multi-slices. In single-slice analysis, carotid arteries were evaluated from a single axial CTA slice of original slice thickness 0.625 mm and slice spacing of 0.4 mm, located at the carotid bifurcation. For multi-slice analysis, images were resampled to 3 mm slice thickness using the OsiriX MD software resampling plugin (Version 10.0.3, Pixmeo SARL, Bernex, Geneva, Switzerland). 14 consecutive slices of the carotid CTA scans were obtained, from three slices below the carotid bifurcation to 10 slices above [17]. These resulted in a total of 14 CTA images per carotid artery, and a final total of 1848 carotid CTA images. A simplified illustration of these methods have been illustrated in Fig. 1.

3. Approaches to carotid artery disease risk stratification

Three approaches to predict carotid artery status were evaluated; (1) the calcium score alone, (2) radiomic features (±calcium score), and (3) a DL approach (CNN).

3.1. Calcium scoring

Carotid calcification quantification was assessed by using the 'calcium scoring' plug-in of OsiriX on unenhanced CT images, as per previous methodology [15]. The total calcium score per artery represents the sum of single calcium scores from all 14 slices of carotid artery, which includes the common and internal carotid arteries, and is

expressed as a score in Agatston units [18].

3.2. Radiomic feature extraction

The open-source Python-coded radiomics package PyRadiomics (PyRad) was used to extract all radiomic features. PyRad enables the extraction of both first order and higher-order radiomic features that capture spatial interrelationships between pixels [19]. PyRad has been effectively used for radiomic feature extraction in both oncology and cardiovascular settings [20]. A total 93 radiomic features were extracted, spanning 7 different feature classes comprising of: (1) first order, (2) shape, (3) GLCM, (4) GLSZM, (5) GLRLM, (6) NGTDM and (7) GLDM. The formal definitions and equations are available from the online documentation at https://pyradiomics.readthedocs.io/.

A previous paper indicated that of the 93 extracted radiomic features, 10 features demonstrated excellent robustness against perturbations and inter-observer variability [17]. We conducted supplementary analyses that compared the predictive performances of including only these 10 "robust" features versus the full 93 features. Explainability analysis (elaborated in a latter section) was also conducted using these "robust" features only.

3.3. Image settings for radiomic feature extraction

Prior to feature extraction, two image pre-processing schemes were applied to the CTA scans: (a) Original (no pre-processing) and (b) Resegmentation. Resegmentation restricts the upper and lower limits of CT Hounsfield units for consideration [21]. A resegmentation filter was applied in the range of 0–200 HUs, reducing the effects of excess macrocalcification and limiting the effect of luminal contrast and perivascular fat that might have been captured due to human error during segmentation of the region of interest [17].

In summary, the following image settings were investigated in the context of radiomic feature extraction:

- Single-slice analysis (at the carotid bifurcation):
  - o Without resegmentation ('single-slice original').
  - o With resegmentation to 0–200 HU ('single-slice resegmented').
- Multi-slice analysis (all 14 slices of the carotid artery):
  - o Without resegmentation ('multi-slice original').
  - o With resegmentation to 0–200 HU ('multi-slice resegmented').

3.4. Deep learning approach

CNNs are specialised DL architectures characterised by the convolution operation which acts as a filter to extract spatially correlated features of an input image, similar to the receptive field of our photoreceptors [22]. In comparison to traditional machine learning algorithms which require the extraction of pre-defined features from images (e.g the calcium score, or other pre-defined radiomic features), CNNs automatically learn which features in the images make the best predictions [14].

Axial CTA slices were cropped to a 30 × 30 pixel image patch centred around the carotid artery to direct the DL model towards the relevant anatomy (Fig. 1). As per standard DL protocol, carotid images were dynamically augmented before input into the models for training. This image altering process adds an additional layer of robustness by preventing excessive overfitting, and encourages the models to focus on more generalisable features that are more likely to be suggestive of carotid pathology.

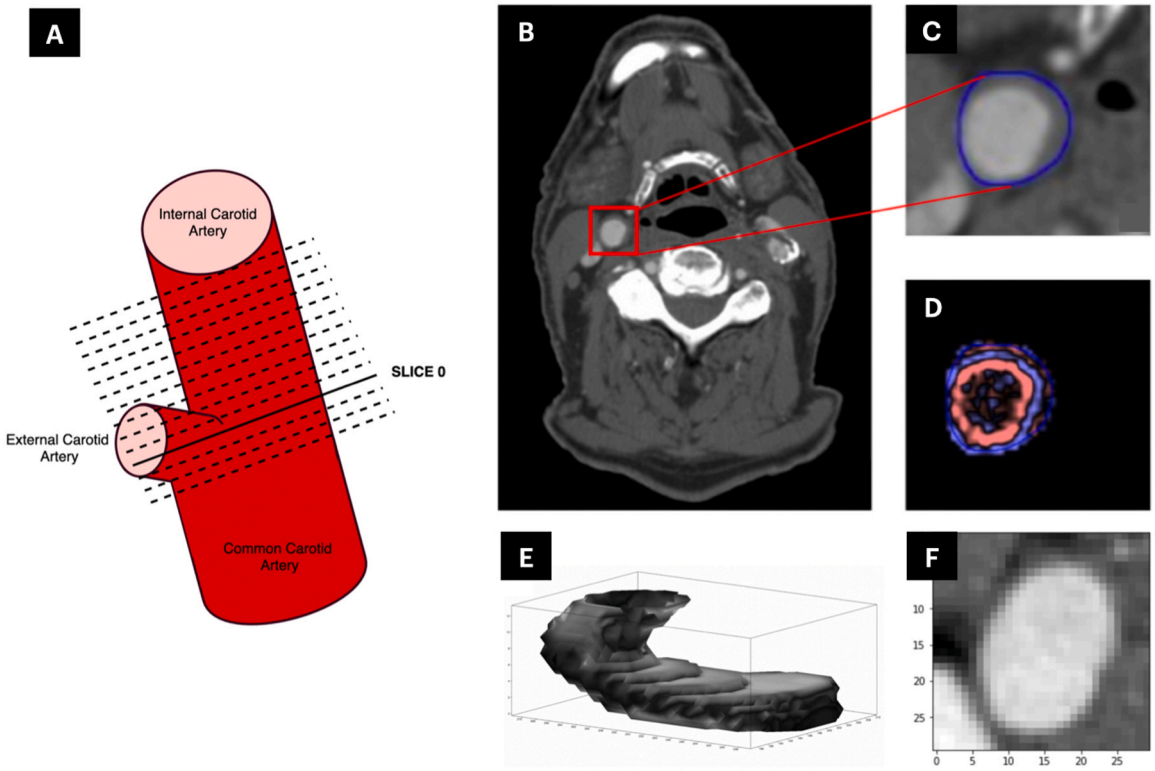

Fig. 1. Conceptual Illustration of delineation of single-slice and multi-slice region of interest for radiomics and deep-learning approaches. 1a: 14 consecutive slices of the carotid CTA scans were obtained, from three slices below the carotid bifurcation to 10 slices above. 1b–d: For each slice, ROIs were manually drawn to encompass the outer vessel wall as illustrated in blue. 1e: The ROI's were automatically propagated for multi-slice analysis along the artery to create a volume, which was then manually adjusted. Multi-slice extraction of radiomic features were applied towards the entire volume. 1f: For deep-learning, a 30 × 30 pixel image patch centred around the manually drawn carotid ROI was cropped to direct the DL models towards relevant anatomy. Multi-slice deep-learning analysis comprised of 2D analysis of each of the 14 slices individually, before obtaining a composite probability.

3.5. Transfer learning

To compensate for the relatively low numbers of images (in a DL context), transfer learning involves adapting existing pre-trained CNN models that have been trained on large generic image datasets [23]. The VGG-16 architecture [24] from the Python Keras library (pre-trained on the ImageNet Large Scale Visual Recognition Challenge with over 1.2 million images) was ultimately selected due to its relatively simple architecture [25].

The VGG-16 architecture consists of 5 blocks of convolutional layers which act as the feature extractors of the DL model [24]. The convolutional layers then feed into the final block of the model that acts as a classifier which assigns weights to the extracted features mapping them to the final output classes [24]. Two architectures of the VGG-16 were used, one preserving the entire original architecture (VGG-16 Original), and one that set the final convolutional layer (block 5) to trainable (VGG-16 Trainable).

4. Machine learning and statistical methods

The three approaches (calcium score vs radiomic features vs DL) were assessed based on performance across three classification tasks: (1) asymptomatic vs symptomatic carotid arteries ('Asx vs Sx'), (2) culprit vs non-culprit carotid arteries ('CC vs NC'), and (3) asymptomatic vs culprit vs non-culprit carotid arteries ('multiclass').

Python version 3.6.7 was used for all analysis [26].

4.1. Machine learning classification models

Traditional machine learning classifiers were trained on (1) the calcium score alone, (2) radiomic features, and (3) a combination of both the calcium score + radiomic features. We considered 8 standard ML benchmarks: K-nearest neighbour, Naive Bayes Classifier, Support Vector Classifier, Decision Tree, Random Forest, Elastic Net regression, Neural Network and Gradient Boost.

For DL, a total of 4 CNN models were trained – 2 simple de-novo models were based on basic "out of the box" CNN architectures which were initiated with randomized weights (LeNet and a Simple CNN), while the latter 2 were built on top of the VGG-16 architectures (VGG-16 Original and VGG-16 Trainable) as per transfer learning.

4.2. Evaluation of predictive performance

The area under the receiving-operating characteristic curve (AUC–ROC, or AUC) was chosen to evaluate the predictive performance of the classification models. For additional robustness, model performance was assessed via a five-fold cross-validation scheme [27]. Data was partitioned into 5 parts, with each part having its turn as the blinded "test" set, whilst the models were trained on the other 4 parts. This method generates a set of 5 final AUC values for each model and is presented as mean AUC (± standard deviation) for all models. Accuracy was also calculated and presented in the Supplementary tables. For illustration, ROC curves for the best models were presented for binary classification tasks, while "composite" confusion matrices (summated predictions of the 5 test sets) are presented for multi-class classification. Predictions were calculated at the optimal cut-point (Youden's Index).

5. Explainability

5.1. Explainability of radiomics-based models: SHAP values

To better comprehend the highest performing classifier algorithms, the SHAP (SHapley Additive exPlanations) approach was used to assess the relative importance of the processed radiomic features [28]. The SHAP approach is a model agnostic approach that employs the "Shapley Value" to assess the relative contribution of each radiomic feature to the decision function of a classifier [28]. ShapVal has been shown to generate consistent accurate reflections of feature for model predictions [29]. We ranked the global importance of the radiomic features based on mean absolute ShapVal as described in previous studies.

5.2. Deep learning explainability: GRAD-CAM visualisation

The Grad-CAM algorithm was developed by Selvaraju et al. in 2017 and stands for gradient- weighted class activation mapping [30]. This algorithm is useful for understanding which parts of a given image contributes to the model's classification decision, and can help to identify the regions of interest where the model is focusing. For "black box" algorithms such as CNNs wherein the extracted features are not user pre-defined, this serves as a additional layer of robustness analysis to ensure that the algorithms are "looking in the right areas".

6. Results

6.1. Overall study characteristics

66 patients were included in this study, comprising 41 symptomatic patients (82 carotids: 41 culprit, 41 non-culprit) and 25 asymptomatic patients (50 asymptomatic carotids). The mean age was 71.4 (SD 9.2) years, of which 75.8 % were male. Supplementary Table 1 summarises the characteristics of all patients (n = 66) and all carotid arteries (n = 132).

6.1.1. Performance of radiomics feature-based approach

Broadly, radiomics models derived from multi-slice images performed better than those derived from the single-slice images. Radiomics models attained the highest AUC in Asx vs Sx classification, followed by multi-class classification, then NC vs CC classification in that order. For multi-slice analysis, image filtering (resegmentation vs original) had little impact on the overall performance, but in single-slice analysis, resegmentation resulted in poorer performance. The overall performance of radiomics models are illustrated in Fig. 2.

Radiomics models attained the highest performance for the binary classification of Sx vs Asx carotids, with both multi-slice original and resegmented radiomic models performing similarly well, achieving mean (SD) AUC values of 0.96 (0.02) and 0.96 (0.03) respectively (Fig. 2**a**). Performance decreased for binary classification of NC vs CC carotids. For this analysis, the multi-slice radiomic models only achieved mean AUC values of 0.75 (0.09), and 0.76 (0.11) respectively. Performance again improved for multi-class classification, with multi-slice original and resegmented models attaining relatively high mean AUC values of 0.84 (0.04) and 0.85 (0.07) (Fig. 2**c**).

Amongst the classifier algorithms, no standout algorithm consistently performed better than its counterparts. Fig. 3 illustrates a comparison of the performance between the commonly high performing classifiers across various settings. Detailed AUC values for each image setting and classifier algorithm for the radiomics feature-based approach can be found in Supplementary Table 2**.** Lastly, as a sub-analysis, classifiers fed only "robust" radiomics features were evaluated and attained largely equivalent performances (Supplementary Table 3**).**

6.1.2. Predictive value of calcium (CAC) score

Table 1 summarises the predictive performance of the calcium score alone, and in conjunction with radiomic features. Overall, the calcium score alone performed poorly in comparison to radiomic models across all three classification tasks. The mean AUC values for calcium were 0.79 (0.08), 0.60 (0.02) and 0.71 (0.07), for Sx vs Asx, NC vs CC, and multiclass classification respectively. The combination of calcium and radiomics did not substantially improve the AUC scores for Sx vs Asx and multiclass classification, though it achieved the highest mean AUC score for binary NC vs CC classification (mean AUC: 0.79 (0.15)) amongst all models.

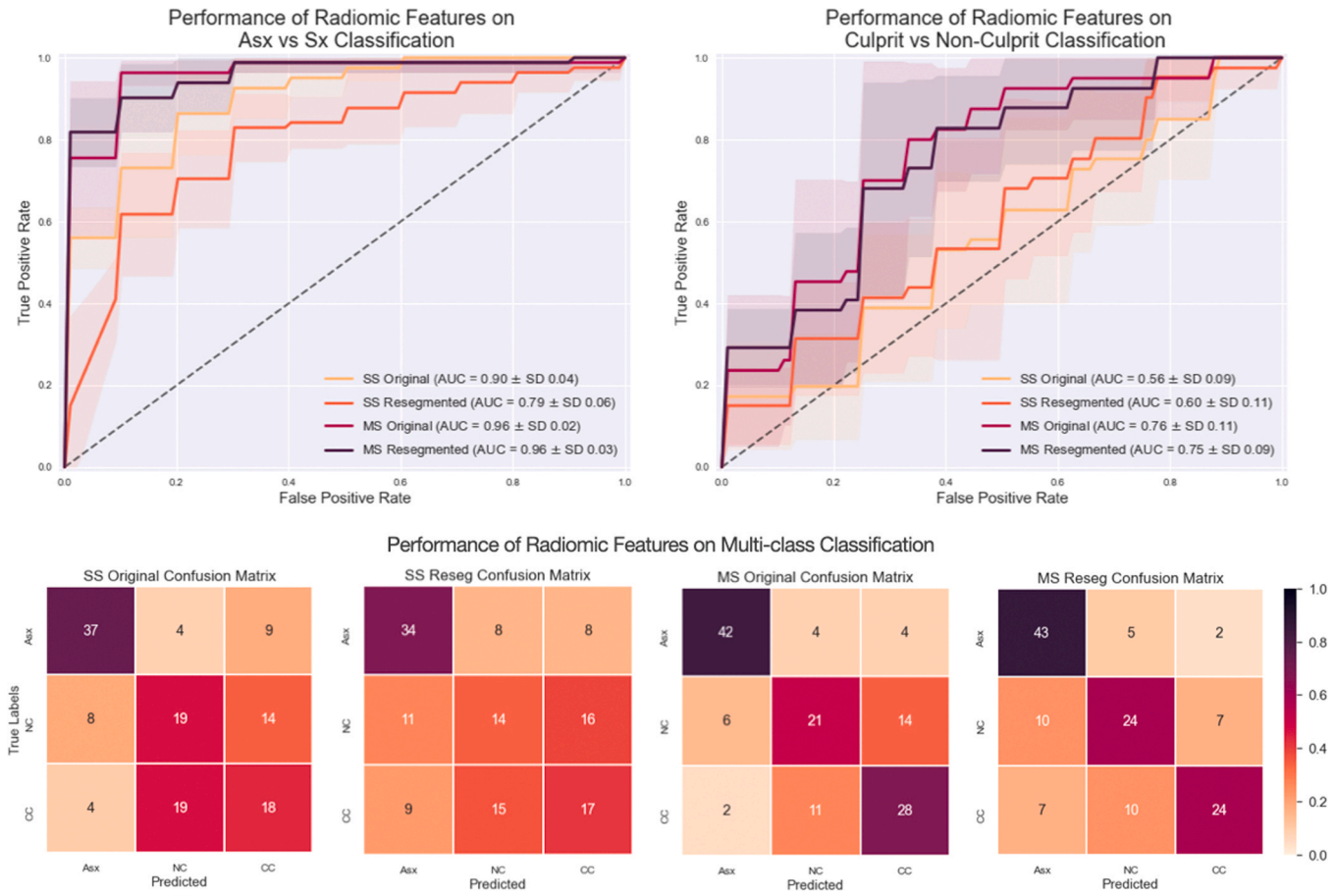

Fig. 2. Overall predictive performance of radiomic models. Fig. 2 illustrates the overall predictive performance of the radiomic models for (2a) Asymptomatic vs Symptomatic arteries, (2b) culprit vs non-culprit arteries, and (2c) multi-class classification (Asymptomatic vs culprit vs non-culprit). The multislice resegmented approach consistently attained the highest predictive performance across all tasks.

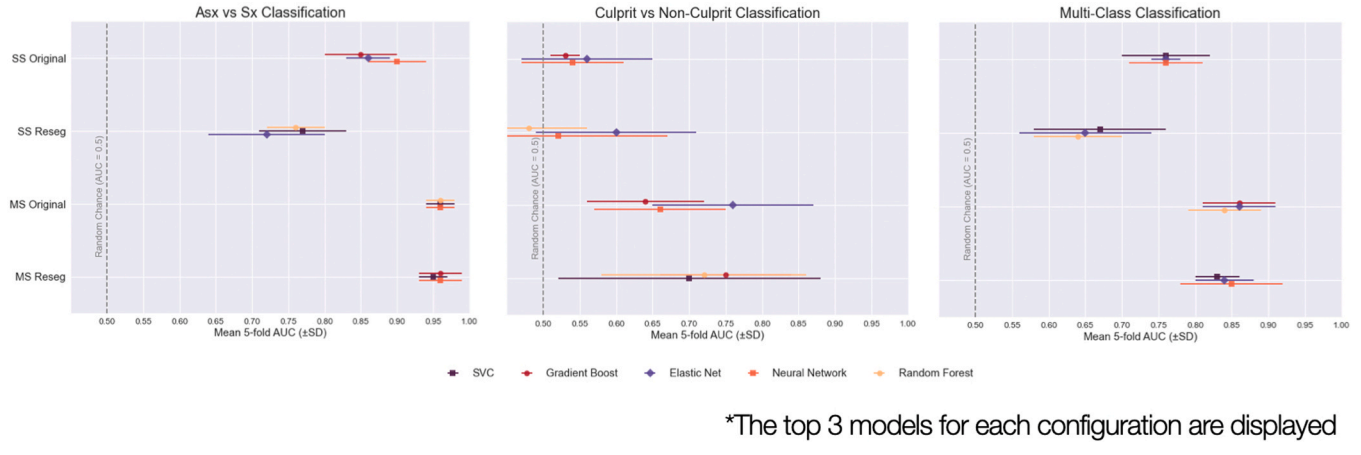

Fig. 3. Comparison of performance achieved by Top Machine Learning Algorithms using the Radiomic Features based approach. Fig. 3 illustrates a comparison of the performance between the more commonly high performing classifiers across the 3 classification tasks. In terms of classifier algorithms, no standout algorithm consistently performed better than its counterparts.

6.1.3. Performance of deep learning (CNN) approach

Similar to the radiomics models, the DL approach likewise attained the highest AUC values for Asx vs Sx classification, followed by multi-class classification and NC vs CC classification. Of the 4 deep learning architectures implemented, the latter 2 architectures based on transfer learning (VGG-16) consistently performed better than their de-novo "out of the box" counterparts. The VGG-16 trainable model consistently offered the highest predictive performance amongst the 4 architectures, with mean AUC values of 0.86 (0.06), 0.67 (0.10) and 0.79 (0.05) for Asx vs Sx, NC vs CC, and multi-class classification tasks respectively. A summary of the overall performance of the DL approach is illustrated in Fig. 4, with a more detailed table available in Supplementary Table 4.

Lastly, Fig. 5 offers an illustrative comparison between the top radiomics, deep learning and calcium models respectively. The

Table 1
Predictive value of calcium score vs radiomic features vs a combination of both.

		Binary: Sx vs Asx			Binary: NC vs CC			Multiclass		
		Calcium Alone	Radiomics Alone	Calcium + Radiomics	Calcium Alone	Radiomics Alone	Calcium + Radiomics	Calcium Alone	Radiomics Alone	Calcium + Radiomics
Top Classifier 1	**AUC**	0.79 (0.08)	0.96 (0.03)	0.97 (0.03)	0.60 (0.02)	0.75 (0.09)	0.79 (0.15)	0.71 (0.07)	0.85 (0.07)	0.85 (0.04)
	Accuracy	0.72 (0.09)	0.89 (0.05)	0.91 (0.03)	0.60 (0.03)	0.67 (0.14)	0.72 (0.13)	0.47 (0.04)	0.68 (0.08)	0.70 (0.05)
Top Classifier 2	**AUC**	0.79 (0.09)	0.96 (0.03)	0.96 (0.02)	0.55 (0.10)	0.72 (0.14)	0.74 (0.18)	0.70 (0.06)	0.84 (0.04)	0.85 (0.04)
	Accuracy	0.64 (0.03)	0.86 (0.05)	0.89 (0.06)	0.50 (0.08)	0.68 (0.10)	0.71 (0.15)	0.48 (0.04)	0.68 (0.08)	0.67 (0.11)
Top Classifier 3	**AUC**	0.77 (0.08)	0.95 (0.02)	0.96 (0.03)	0.54 (0.06)	0.71 (0.17)	0.72 (0.17)	0.70 (0.07)	0.83 (0.03)	0.85 (0.05)
	Accuracy	0.67 (0.03)	0.87 (0.04)	0.86 (0.05)	0.50 (0.08)	0.67 (0.15)	0.65 (0.19)	0.46 (0.04)	0.70 (0.07)	0.73 (0.04)

radiomics models (multi-slice resegmented approach) offered the highest predictive performance for all three classification tasks, followed by the deep learning (VGG-16 trainable) approach, and lastly the calcium score alone.

6.1.4. *Explainability*

For explainability, Table 2 presents the radiomic features as ranked by their mean absolute SHAP values. The "key" features as identified by SHAP analysis were often quite consistent across the various classifiers and classifier tasks. "GLRM: Long Run High Gray Level Emphasis" was consistently a key feature for all Sx vs Asx, NC vs CC and Multiclass classification tasks. "GLDM: Dependence Variance" and "GLDM: Large Dependence High Gray Level Emphasis" were key features for Sx vs Asx and Multiclass classification, while "GLCM: Difference Variance" was a key feature for NC vs CC and Multiclass classification. These features also consistently had the highest absolute beta-coefficients in the Logistic Regression models (Supplementary Table 4**).**

Similarly, for the DL models, the GRAD-CAM visualisations of the "key image areas" are illustrated in Fig. 6. The GRAD-CAM mappings for the de-Novo CNN architectures appear closely centred around the carotid artery walls and lumens, whilst the mappings for the VGG-16 transfer architectures highlight a more diffused hot region covering areas within the central lumen but also including peri-vascular areas.

7. Discussion

7.1. *Overall key findings*

To our knowledge, this is the first systematic "proof of principle" study that evaluates the ability of radiomic features and DL methods to identify carotid artery disease from carotid CT angiography images. We investigated the performance of these novel image analysis techniques across 3 classification tasks: Asymptomatic vs Symptomatic, Culprit vs non-Culprit, and Multiclass (Asymptomatic vs Culprit vs non-Culprit). We also demonstrated superior classification performance compared to the traditional carotid calcium score. Lastly, we applied SHAP and GRAD-CAM analysis for explainability of these machine learning models.

We report the following key findings: (1) both radiomic features and deep learning methods offered better predictive accuracy of CAD than the calcium score, (2) radiomic features attained the highest performance within our dataset, (3) a multi-slice approach for radiomic feature extraction was the most optimum setting, and (4) there may be utility for a combination of novel and traditional features (ie radiomics + calcium score). Together, these findings highlight the potential of using radiomic and DL methods in the clinical assessment of carotid artery disease. As CT angiography already forms part of the workflow in ischaemic cerebrovascular management, there is ripe potential for the integration of these novel imaging biomarkers into the clinical practice for better risk stratification and tailoring of patient management[31].

Previous studies have investigated the use of other CTA derived metrics to evaluate symptomatic carotid artery disease. Gupta et al. investigated the discriminative ability of CTA plaque thickness [32], while Magge et al. [33] and Motoyama et al. [34] both assessed the utility of defined morphological plaque characteristics (i.e. wall thickness, fibrous cap thickness, lipid-rich necrotic core). Lastly, a pilot study by Zaccagna et al. highlighted the potential of texture analysis in identifying vulnerable patients in stroke and TIA [11]. Our results further reinforce the potential of the additional information that can be mined from carotid CT angiograms.

7.1.1. *Radiomics models*

The radiomics models offered the greatest performance across all three classification tasks. Radiomic features extracted from multi-slice settings performed considerably better than their single-slice counterparts. These radiomics models were best able to differentiate carotid

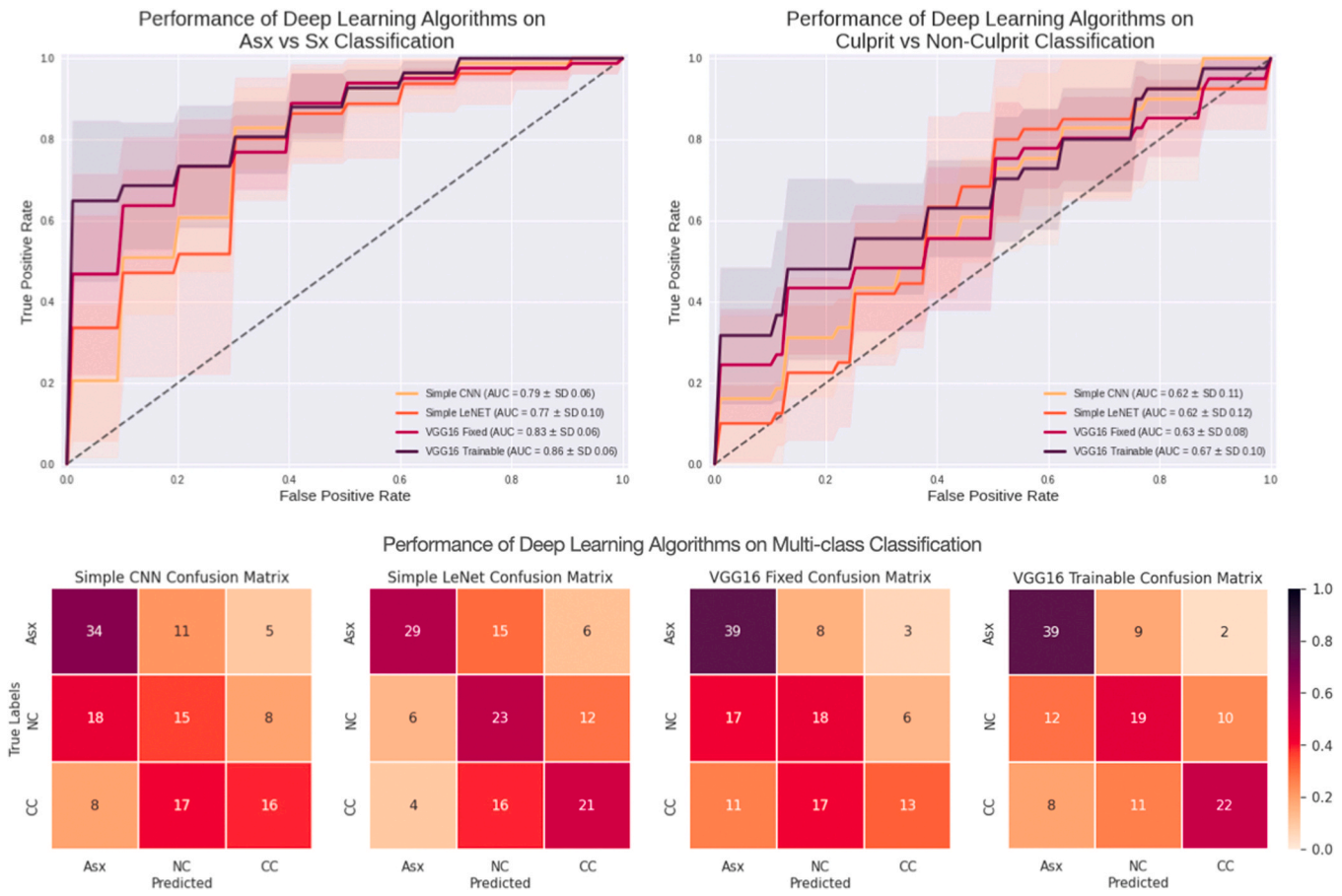

Fig. 4. Overall Predictive Performance of Deep Learning (Convolutional Neural Network) Approach. Fig. 4 illustrates the overall predictive performance of the various deep learning convolutional neural network models for (4a) Asymptomatic vs Symptomatic arteries, (4b) culprit vs non-culprit arteries, and (4c) multi-class classification (Asymptomatic vs culprit vs non-culprit). The VGG16 trainable model (transfer learning approach) consistently attained the highest predictive performance.

arteries between symptomatic patients and asymptomatic patients, and encountered more difficulties differentiating culprit versus non-culprit carotid arteries. These may all reflect differences in the presence of carotid atherosclerotic plaques, carotid calcium burden and stenosis [1, 17].

7.1.2. *Calcium score utility*

As a univariable predictor, the carotid calcium score had the poorest predictive performance in comparison to other approaches. However, a combination of radiomics + calcium did appear to marginally achieve the highest predictive performance for Culprit vs Non-Culprit classification (0.79 ± 0.14), though we are wary of the large standard deviation observed in the set of AUC values. This result suggests that carotid calcium and the information captured by radiomic features (with resegmentation) may contain complementary information that can be exploited with multivariable models.

7.1.3. *Deep learning*

Our results also demonstrated a potential utility for deep-learning (CNNs) methods in the assessment of carotid artery disease. We note however, that whilst the DL models outperformed the calcium score, they performed more poorly than the radiomics models. The performance of DL models scale with sample size, wherein they often outperform traditional models when trained on large sample sets. Due to our relatively small sample size (within the context of DL), it is likely the DL models were unable to achieve optimal optimization of their large number of parameters. This observation was reinforced when the application of transfer learning (the VGG models) resulted in considerable improvement over the "out of the box" architectures. Lastly, we note that the DL models were given a 'harder' task, wherein radiomic models were fed features bounded within the specific carotid region of interest, whilst the entire 30×30 image was fed into the DL models. Ultimately, our results indicate even though deep learning is potentially a viable method for learning new features from carotid imaging data, 1) more data and 2) refinement of the deep learning methods is still needed. For centres looking to develop personalised image analysis models tailored to regional based imaging modalities and populations, our results suggest it might be more pragmatic to adopt radiomic feature analysis over DL models in the absence of larger local training datasets.

7.1.4. *Explainability*

SHAP-value analysis of the radiomic feature approach revealed consistent patterns in the identification of high value radiomic features. "GLRM: Long Run High Grey Level Emphasis", "GLDM: Dependence Variance", "GLDM: Large Dependence High Grey Level Emphasis" and "GLCM: Difference Variance" were all features deemed "important" to the decision function of the classifier functions. These are higher-order radiomic features that considered the spatial interrelationships between pixels, which reflect a degree of complexity in the information that the radiomic models are extracting.

For GRAD-CAM visualisations of the DL models, GRAD-CAM visualisations were focused closely on the contours of the carotid arteries in the simpler models, whilst displaying a more diffuse region of focus in the VGG-16 models. These differences are likely due to differences in the depth of architecture between the VGG-16 vs simpler DL models, which

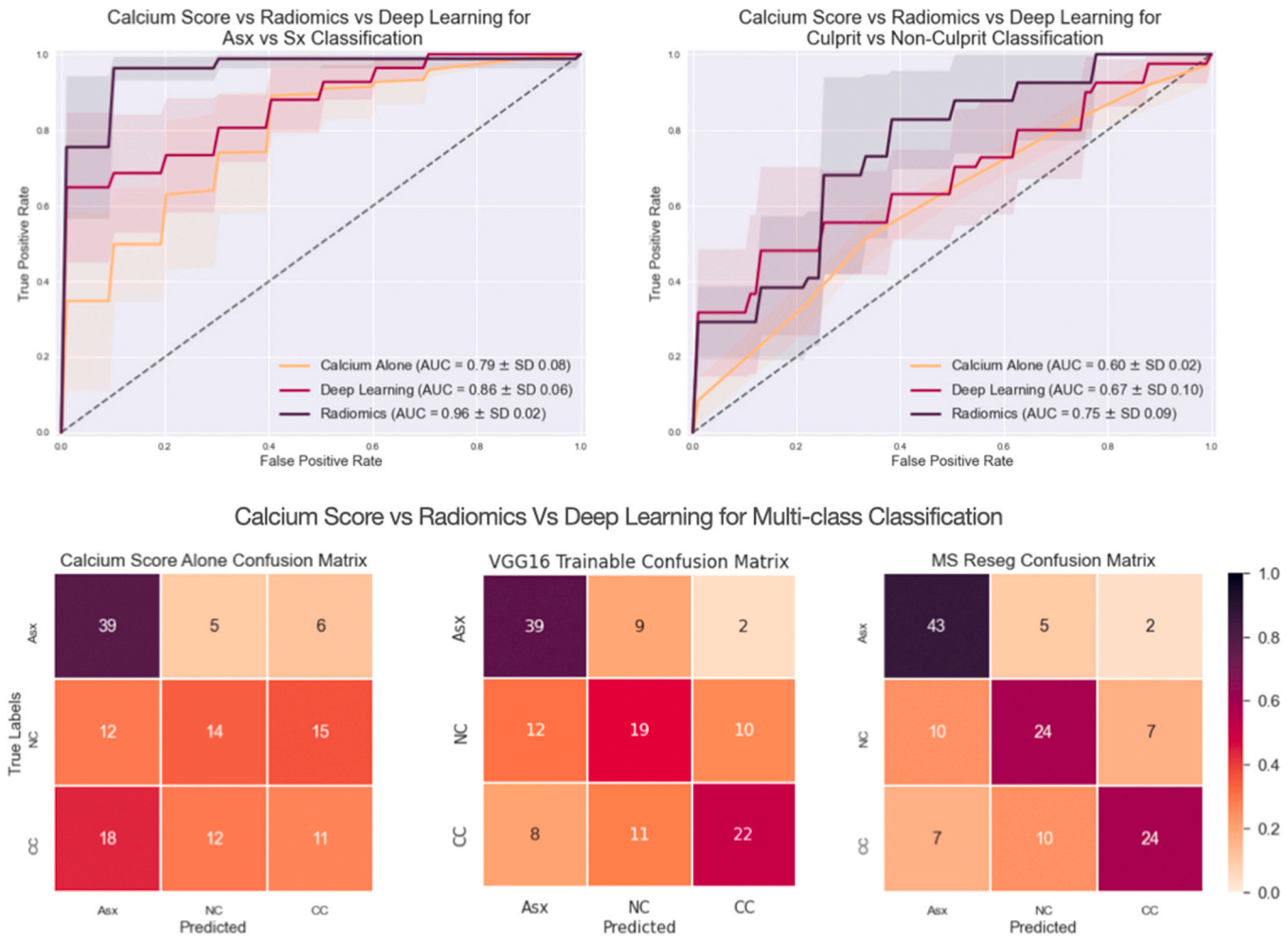

Fig. 5. Comparison of Predictive Performance (Calcium Scores vs Radiomic Features Vs Deep Learning). Fig. 5 illustrates the overall predictive performance of calcium vs radiomics vs deep learning for (5a) Asymptomatic vs Symptomatic arteries, (5b) culprit vs non-culprit arteries, and (5c) multi-class classification (Asymptomatic vs culprit vs non-culprit). The radiomics model (multi-slice resegmented approach) attained the highest predictive performance across all three tasks.

likely resulted in heavier upsampling and interpolation of the VGG-16 GRAD-CAM visualisations. It was nonetheless reassuring that the key regions for all models were all closely centred around the carotid arteries.

7.1.5. Limitations

This study has a few limitations. First, we note the comparatively small sample size in the context of deep-learning analysis, as discussed in above sections. Second, all images were acquired using the same scanner and same scanning protocols in a single centre, which limit the generalisability of our findings. Third, this was a retrospective study, and our imaging dataset only captured information from culprit carotid arteries after plaque rupture had occurred. Prospective studies of at-risk patients prior to recurrent stroke events are warranted. Fourth, we were unable to assess carotid stenosis as an unbiased "predictive marker" due to the nature of the inclusion criteria for the imaging studies from which the pooled dataset was derived. The ICARUSS and CHAI studies for example, mandated $\geq$ 50 % and $\geq$ 30 % stenosis in the culprit carotid arteries. Consequently, most symptomatic carotid arteries already had a significant degree of stenosis. Lastly, the primary objective of this study was to evaluate the utility of radiomics and DL approaches as a "proof of principle", rather than developing a definitive radiomics signature or CNN architecture. As such, the default configurations for most machine learning classifiers were used, without extensive hyperparameter tuning – generally an accepted approach as per previous studies [35]. Whilst these models may not yet be robust enough for implementation in current clinical practice, our results pave the way for future studies to further optimize and validate these parameters and workflows for future clinical implementation.

8. Conclusion

Our "proof of principle" study demonstrates the feasibility of radiomics and deep learning approaches in the identification of carotid artery disease. This highlights the potential clinical utility of these novels techniques in extracting quantitative information from CTA images beyond calcification. Future work is warranted to explore methods to optimise carotid CTA workflows, and the acquisition of prospective data for clinical validation will be crucial to support the transition of these techniques into the clinical practice of cardiology and stroke medicine. This will ultimately facilitate better stroke risk stratification and a shift towards personalised patient management.

CRediT authorship contribution statement

Elizabeth PV Le: Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mark Wong**: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation. **Leonardo Rundo**: Writing – review & editing, Supervision, Methodology, Investigation, Data curation, Conceptualization. **Jason M Tarkin**: Writing – review & editing, Resources, Methodology, Investigation, Data curation, Conceptualization.

Table 2
Explainability of radiomic features as assessed by SHAP values.

Feature	Binary: Sx vs Asx			Binary: NC vs CC			Multiclass		
	Top Classifier 1	Top Classifier 2	Top Classifier 3	Top Classifier 1	Top Classifier 2	Top Classifier 3	Top Classifier 1	Top Classifier 2	Top Classifier 3
First Order: Mean Absolute Deviation',				4				3	
GLCM: Difference Variance',					1	3	2	2	2
'GLCM: Joint Energy',									
'GLDM: Dependence Variance',	1	1					1	1	1
'GLDM: Large Dependence High Gray Level Emphasis',	2	3	1			2	4		3
'GLRLM: Long Run High Gray Level Emphasis',	3	4	3	1	2	1	3	4	4
'GLRLM: Gray Level NonUniformity Normalized',				2		4			
'GLSZM: Gray Level NonUniformity',	4		4		3				
'GLSZM: Large Area High Gray Level Emphasis',		2	2	3					
'GLSZM: Size Zone NonUniformity'					4				

Legend: Radiomic features are ranked according to the magnitude of the absolute SHAP value, and color coded by importance, with the darkest cells indicating the most “influential” feature

Radiomic features are ranked according to the magnitude of the absolute SHAP value, and color coded by importance, with the darkest cells indicating the most “influential” feature.

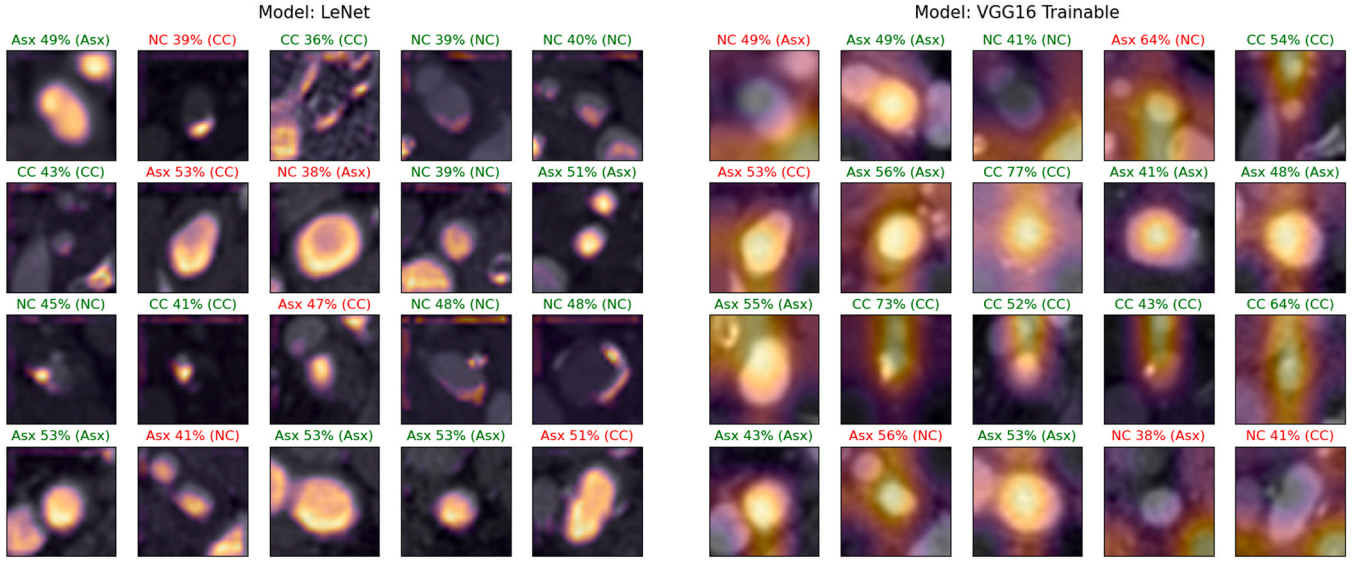

Fig. 6. Explainability of Deep Learning Algorithms: GRAD-CAM Visualisation. The GRAD-CAM algorithm was applied to both the LeNet and VGG16 trainable deep learning models to identify the regions of interest that contributes to the model classification decision.

Nicholas R Evans: Writing – review & editing, Resources, Data curation, Conceptualization. **Jonathan R Weir-McCall**: Writing – review & editing, Resources, Methodology, Investigation, Data curation, Conceptualization. **Mohammed M Chowdhury**: Writing – review & editing, Supervision, Resources, Investigation, Data curation, Conceptualization. **Patrick A Coughlin**: Writing – review & editing, Resources, Methodology, Data curation, Conceptualization. **Holly Pavey**: Writing – review & editing, Resources, Data curation, Conceptualization. **Fulvio Zaccagna**: Writing – review & editing, Supervision, Methodology, Investigation, Data curation, Conceptualization. **Chris Wall**: Writing – review & editing, Investigation, Data curation, Conceptualization. **Rouchelle Sriranjan**: Writing – review & editing, Resources, Data curation, Conceptualization. **Andrej Corovic**: Writing – review & editing, Supervision, Data curation, Conceptualization. **Yuan Huang**: Writing – review & editing, Supervision, Methodology, Data curation, Conceptualization. **Elizabeth A Warburton**: Writing – review & editing, Supervision, Data curation, Conceptualization. **Evis Sala**: Writing – review & editing, Resources, Data curation, Conceptualization. **Michael Roberts**: Writing – review & editing, Supervision, Software, Methodology, Investigation, Data curation, Conceptualization. **Carola-Bibiane Schönlieb**: Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Data curation, Conceptualization. **James HF Rudd**: Writing – review & editing, Supervision, Resources, Project administration, Methodology,

Investigation, Funding acquisition, Data curation, Conceptualization.

Ethical Statements

This retrospective study analysed unenhanced CT and CTA images pooled from three previous observational research vascular imaging datasets (ICARUSS, VISION and CHAI) originating from a single institution (Addenbrooke's Hospital, Cambridge, UK) between 2011 and 2016. Full details on the methodology, ethical approvals and inclusion and exclusion criteria have been published previously on:

N.R. Evans, J.M. Tarkin, M.M. Chowdhury, E.P.V. Le, P.A. Coughlin, J.H.F. Rudd, E.A. Warburton, Dual-Tracer Positron-Emission Tomography for Identification of Culprit Carotid Plaques and Pathophysiology In Vivo, Circ Cardiovasc Imaging 13(3) (2020) e009539.

J.M. Tarkin, F.R. Joshi, N.R. Evans, M.M. Chowdhury, N.L. Figg, A.V. Shah, L.T. Starks, A. Martin-Garrido, R. Manavaki, E. Yu, R.E. Kuc, L. Grassi, R. Kreuzhuber, M.A. Kostadima, M. Frontini, P.J. Kirkpatrick, P. A. Coughlin, D. Gopalan, T.D. Fryer, J.R. Buscombe, A.M. Groves, W.H. Ouwehand, M.R. Bennett, E.A. Warburton, A.P. Davenport, J.H. Rudd, Detection of Atherosclerotic Inflammation by (68)Ga-DOTATATE PET Compared to [(18)F]FDG PET Imaging, J Am Coll Cardiol 69(14) (2017) 1774–1791.

Declaration of Competing Interest

The authors declare no competing interests.

Funding and Acknowledgements

JHFR is part-supported by the NIHR Cambridge Biomedical Research Centre, the British Heart Foundation, HEFCE, the EPSRC and the Wellcome Trust. JRWM is supported by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014). JMT is supported by the Wellcome Trust (211100/Z/18/Z) and the Cambridge British Heart Foundation Centre of Research Excellence (18/1/34212). AC is supported by a British Heart Foundation Clinical Research Training Fellowship, FS/CRTF/20/24035. NRE is supported by The Stroke Association (SA-SCL-MED-22/100006) and NIHR Cambridge Biomedical Research Centre (NIHR203312).

The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejro.2024.100594.

References

[1] M.M. Mughal, M.K. Khan, J.K. DeMarco, A. Majid, F. Shamoun, G.S. Abela, Symptomatic and asymptomatic carotid artery plaque, Expert Rev. Cardiovasc. Ther. 9 (10) (2011) 1315–1330.
[2] A.Y. Yu, S.B. Coutts, Stroke: risk assessment to prevent recurrence after mild stroke or TIA, Nat. Rev. Neurol. 11 (3) (2015) 131–133.
[3] H.J. Barnett, D.W. Taylor, M. Eliasziw, A.J. Fox, G.G. Ferguson, R.B. Haynes, R. N. Rankin, G.P. Clagett, V.C. Hachinski, D.L. Sackett, K.E. Thorpe, H.E. Meldrum, J. D. Spence, Benefit of carotid endarterectomy in patients with symptomatic moderate or severe stenosis. North American Symptomatic Carotid Endarterectomy Trial Collaborators, N. Engl. J. Med. 339 (20) (1998) 1415–1425.
[4] A. Rerkasem, S. Orrapin, D.P. Howard, K. Rerkasem, Carotid endarterectomy for symptomatic carotid stenosis, Cochrane Database Syst. Rev. 9 (9) (2020) Cd001081.
[5] M.J. Bom, D.J. van der Heijden, E. Kedhi, J. van der Heyden, M. Meuwissen, P. Knaapen, S.A.J. Timmer, N. van Royen, Early detection and treatment of the vulnerable coronary plaque: can we prevent acute coronary syndromes? Circ. Cardiovasc. Imaging 10 (5) (2017).
[6] D. Bos, B. Arshi, Q.J.A. van den Bouwhuijsen, M.K. Ikram, M. Selwaness, M. W. Vernooij, M. Kavousi, A. van der Lugt, Atherosclerotic carotid plaque composition and incident stroke and coronary events, J. Am. Coll. Cardiol. 77 (11) (2021) 1426–1435.
[7] R. Mortimer, S. Nachiappan, D.C. Howlett, Carotid artery stenosis screening: where are we now? Br. J. Radiol. 91 (1090) (2018) 20170380.
[8] R.M. Kwee, Systematic review on the association between calcification in carotid plaques and clinical ischemic symptoms, J. Vasc. Surg. 51 (4) (2010) 1015–1025.
[9] M. Ahmed, R. McPherson, A. Abruzzo, S.E. Thomas, V.R. Gorantla, Carotid artery calcification: what we know so far, Cureus 13 (10) (2021) e18938.
[10] R. Cau, A. Flanders, L. Mannelli, C. Politi, G. Faa, J.S. Suri, L. Saba, Artificial intelligence in computed tomography plaque characterization: a review, Eur. J. Radiol. 140 (2021) 109767.
[11] F. Zaccagna, B. Ganeshan, M. Arca, M. Rengo, A. Napoli, L. Rundo, A.M. Groves, A. Laghi, I. Carbone, L.J. Menezes, CT texture-based radiomics analysis of carotid arteries identifies vulnerable patients: a preliminary outcome study, Neuroradiology 63 (7) (2021) 1043–1052.
[12] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[13] G. Litjens, F. Ciompi, J.M. Wolterink, B.D. de Vos, T. Leiner, J. Teuwen, I. Išgum, State-of-the-art deep learning in cardiovascular image analysis, JACC Cardiovasc. Imaging 12 (8 Pt 1) (2019) 1549–1565.
[14] M.P. McBee, O.A. Awan, A.T. Colucci, C.W. Ghobadi, N. Kadom, A.P. Kansagra, S. Tridandapani, W.F. Auffermann, Deep learning in radiology, Acad. Radiol. 25 (11) (2018) 1472–1480.
[15] N.R. Evans, J.M. Tarkin, M.M. Chowdhury, E.P.V. Le, P.A. Coughlin, J.H.F. Rudd, E.A. Warburton, Dual-tracer positron-emission tomography for identification of culprit carotid plaques and pathophysiology in vivo, Circ. Cardiovasc. Imaging 13 (3) (2020) e009539.
[16] J.M. Tarkin, F.R. Joshi, N.R. Evans, M.M. Chowdhury, N.L. Figg, A.V. Shah, L. T. Starks, A. Martin-Garrido, R. Manavaki, E. Yu, R.E. Kuc, L. Grassi, R. Kreuzhuber, M.A. Kostadima, M. Frontini, P.J. Kirkpatrick, P.A. Coughlin, D. Gopalan, T.D. Fryer, J.R. Buscombe, A.M. Groves, W.H. Ouwehand, M. R. Bennett, E.A. Warburton, A.P. Davenport, J.H. Rudd, Detection of atherosclerotic inflammation by (68)Ga-DOTATATE PET compared to [(18)F]FDG PET imaging, J. Am. Coll. Cardiol. 69 (14) (2017) 1774–1791.
[17] E.P.V. Le, L. Rundo, J.M. Tarkin, N.R. Evans, M.M. Chowdhury, P.A. Coughlin, H. Pavey, C. Wall, F. Zaccagna, F.A. Gallagher, Y. Huang, R. Sriranjan, A. Le, J. R. Weir-McCall, M. Roberts, F.J. Gilbert, E.A. Warburton, C.B. Schönlieb, E. Sala, J. H.F. Rudd, Assessing robustness of carotid artery CT angiography radiomics in the identification of culprit lesions in cerebrovascular events, Sci. Rep. 11 (1) (2021) 3499.
[18] A.S. Agatston, W.R. Janowitz, F.J. Hildner, N.R. Zusmer, M. Viamonte, Jr, R. Detrano, Quantification of coronary artery calcium using ultrafast computed tomography, J. Am. Coll. Cardiol. 15 (4) (1990) 827–832.
[19] J.J.M. van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R.G.H. Beets-Tan, J.C. Fillion-Robin, S. Pieper, H. Aerts, Computational radiomics system to decode the radiographic phenotype, Cancer Res. 77 (21) (2017) e104–e107.
[20] M. Yang, Y. Ren, Y. She, D. Xie, X. Sun, J. Shi, G. Zhao, C. Chen, Imaging phenotype using radiomics to predict dry pleural dissemination in non-small cell lung cancer, Ann. Transl. Med. 7 (12) (2019) 259.
[21] A. Ibrahim, M. Vallières, H. Woodruff, S. Primakov, M. Beheshti, S. Keek, T. Refaee, S. Sanduleanu, S. Walsh, O. Morin, P. Lambin, R. Hustinx, F.M. Mottaghy, Radiomics analysis for clinical decision support in nuclear medicine, Semin. Nucl. Med. 49 (5) (2019) 438–449.
[22] G. Chartrand, P.M. Cheng, E. Vorontsov, M. Drozdzal, S. Turcotte, C.J. Pal, S. Kadoury, A. Tang, Deep learning: a primer for radiologists, Radiographics 37 (7) (2017) 2113–2131.
[23] K. Weiss, T.M. Khoshgoftaar, D. Wang, A survey of transfer learning, J. Big Data 3 (1) (2016) 1–40.
[24] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556, 2014.
[25] E. Baum, F. Wilczek, Supervised learning of probability distributions by neural networks, Neural Inf. Process. Syst., 1987.
[26] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[27] P. Refaeilzadeh, L. Tang, H. Liu, Cross-validation, Encycl. Database Syst. 5 (2009) 532–538.
[28] S.M. Lundberg, G. Erion, H. Chen, A. DeGrave, J.M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.I. Lee, From local explanations to global understanding with explainable AI for trees, Nat. Mach. Intell. 2 (1) (2020) 56–67.
[29] S.M. Lundberg, B. Nair, M.S. Vavilala, M. Horibe, M.J. Eisses, T. Adams, D. E. Liston, D.K. Low, S.F. Newman, J. Kim, S.I. Lee, Explainable machine-learning predictions for the prevention of hypoxaemia during surgery, Nat. Biomed. Eng. 2 (10) (2018) 749–760.
[30] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: visual explanations from deep networks via gradient-based localization, Int. J. Comput. Vis. 128 (2) (2020) 336–359.
[31] P. Lambin, R.T.H. Leijenaar, T.M. Deist, J. Peerlings, E.E.C. de Jong, J. van Timmeren, S. Sanduleanu, R. Larue, A.J.G. Even, A. Jochems, Y. van Wijk, H. Woodruff, J. van Soest, T. Lustberg, E. Roelofs, W. van Elmpt, A. Dekker, F. M. Mottaghy, J.E. Wildberger, S. Walsh, Radiomics: the bridge between medical imaging and personalized medicine, Nat. Rev. Clin. Oncol. 14 (12) (2017) 749–762.
[32] A. Gupta, H. Baradaran, H. Kamel, A. Pandya, A. Mangla, A. Dunning, R. S. Marshall, P.C. Sanelli, Evaluation of computed tomography angiography plaque thickness measurements in high-grade carotid artery stenosis, Stroke 45 (3) (2014) 740–745.

[33] R. Magge, B.C. Lau, B.P. Soares, S. Fischette, S. Arora, E. Tong, S. Cheng, M. Wintermark, Clinical risk factors and CT imaging features of carotid atherosclerotic plaques as predictors of new incident carotid ischemic stroke: a retrospective cohort study, AJNR Am. J. Neuroradiol. 34 (2) (2013) 402–409.
[34] S. Motoyama, H. Ito, M. Sarai, T. Kondo, H. Kawai, Y. Nagahara, H. Harigaya, S. Kan, H. Anno, H. Takahashi, H. Naruse, J. Ishii, H. Hecht, L.J. Shaw, Y. Ozaki, J. Narula, Plaque characterization by coronary computed tomography angiography and the likelihood of acute coronary events in mid-term follow-up, J. Am. Coll. Cardiol. 66 (4) (2015) 337–346.
[35] A. Arcuri, G. Fraser, Parameter tuning or default values? An empirical investigation in search-based software engineering, Empir. Softw. Eng. 18 (3) (2013) 594–623.