EuroIntervention 2021;17:41-50 published online jaa February 2021

Artificial intelligence and optical coherence tomography for the automatic characterisation of human atherosclerotic plaques

Miao Chu[1], BSc; Haibo Jia[2], MD, PhD; Juan Luis Gutiérrez-Chico[3], MD, PhD; Akiko Maehara[4,5], MD, PhD; Ziad A. Ali[4,5], MD, DPhil; Xiaoling Zeng[6], MD; Luping He[2], MD; Chen Zhao[2], MD; Mitsuaki Matsumura[5], BS; Peng Wu[1], BSc; Ming Zeng[2], MD; Takashi Kubo[7], MD, PhD; Bo Xu[8], MBBS; Lianglong Chen[6], MD, PhD; Bo Yu[2], MD, PhD; Gary S. Mintz[5], MD, PhD; William Wijns[9], MD, PhD; Niels R. Holm[10], MD; Shengxian Tu[1,6]*, PhD

1. School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China; 2. Department of Cardiology, 2nd Affiliated Hospital of Harbin Medical University, Harbin, China; 3. Cardiology Department, Campo de Gibraltar Health Trust, Algeciras, Spain; 4. Center for Interventional Vascular Therapy, Division of Cardiology, Presbyterian Hospital and Columbia University, New York, NY, USA; 5. Cardiovascular Research Foundation, New York, NY, USA; 6. Department of Cardiology, Fujian Medical University Union Hospital, Fuzhou, Fujian, China; 7. Department of Cardiovascular Medicine, Wakayama Medical University, Wakayama, Japan; 8. Fu Wai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences, Beijing, China; 9. The Lambe Institute for Translational Medicine and CURAM, National University of Ireland Galway, Galway, Ireland; 10. Department of Cardiology, Aarhus University Hospital, Aarhus, Denmark

This paper also includes supplementary data published online at: https://eurointervention.pcronline.com/doi/10.4244/EIJ-D-20-01355

KEYWORDS

- intravascular ultrasound
- optical coherence tomography
- stable angina

Abstract

Background: Intravascular optical coherence tomography (IVOCT) enables detailed plaque characterisation *in vivo*, but visual assessment is time-consuming and subjective.

Aims: This study aimed to develop and validate an automatic framework for IVOCT plaque characterisation using artificial intelligence (AI).

Methods: IVOCT pullbacks from five international centres were analysed in a core lab, annotating basic plaque components, inflammatory markers and other structures. A deep convolutional network with encoding-decoding architecture and pseudo-3D input was developed and trained using hybrid loss. The proposed network was integrated into commercial software to be externally validated on additional IVOCT pullbacks from three international core labs, taking the consensus among core labs as reference.

Results: Annotated images from 509 pullbacks (391 patients) were divided into 10,517 and 1,156 cross-sections for the training and testing data sets, respectively. The Dice coefficient of the model was 0.906 for fibrous plaque, 0.848 for calcium and 0.772 for lipid in the testing data set. Excellent agreement in plaque burden quantification was observed between the model and manual measurements (R2=0.98). In the external validation, the software correctly identified 518 out of 598 plaque regions from 300 IVOCT cross-sections, with a diagnostic accuracy of 97.6% (95% CI: 93.4-99.3%) in fibrous plaque, 90.5% (95% CI: 85.2-94.1%) in lipid and 88.5% (95% CI: 82.4-92.7%) in calcium. The median time required for analysis was 21.4 (18.6-25.0) seconds per pullback.

Conclusions: A novel AI framework for automatic plaque characterisation in IVOCT was developed, providing excellent diagnostic accuracy in both internal and external validation. This model might reduce subjectivity in image interpretation and facilitate IVOCT quantification of plaque composition, with potential applications in research and IVOCT-guided PCI.

**Corresponding author: Room 123, Med-X Research Institute, Shanghai Jiao Tong University, No. 1954, Hua Shan Road, Shanghai, 200030, China. E-mail: sxtu@sjtu.edu.cn*

SUBMITTED ON 30/11/2020 - REVISION RECEIVED ON 1st 22/12/2020 / 2nd 22/01/2021 / 3rd 27/01/2021 - ACCEPTED ON 28/01/2021

DOI: 10.4244/EIJ-D-20-01355

Abbreviations

AI	artificial intelligence
CNN	convolutional neural network
DL	deep learning
EEL	external elastic lamina
IEL	internal elastic lamina
IVOCT	intravascular optical coherence tomography
IVUS	intravascular ultrasound
PB	plaque burden
TCFA	thin-cap fibroatheroma

Introduction

Intravascular optical coherence tomography (IVOCT) enables detailed *in vivo* visualisation of atherosclerotic plaque with high resolution and accurate characterisation of different types of tissue composition[1]. This information is instrumental for plaque-specific lesion preparation and for appropriate selection of the therapeutic strategy during IVOCT-guided intervention[2,3]. However, plaque characterisation is currently challenging, time-consuming, difficult to systematise, based on subjective interpretation of the operators and largely dependent on their expertise, thus posing problems for reproducibility.

Artificial intelligence (AI) might be useful to standardise and automate plaque characterisation in interpreting IVOCT images. Deep learning (DL) is a data-driven AI algorithm to extract regular patterns from the observation of data by training a designed model. After training, the model learns features that can be used to make predictions in unexplored data sets. Convolutional neural network (CNN) is a specific architecture in DL that has excellent performance in image processing. Several studies have attempted automatic characterisation of atherosclerotic plaques in coronary arteries using CNN, following different approaches[4-6].

The aim of this study was to develop a deep convolutional network for comprehensive plaque characterisation, trained on a large diverse data set of IVOCT images comprising varied anatomic morphologies and clinical scenarios, to provide both qualitative characterisation and quantification of plaque components for clinical use.

Editorial, see page 18

Methods

STUDY DESIGN AND STUDY POPULATION

This was a retrospective, *post hoc*, multicentre, international study to appraise the accuracy and reproducibility of a novel diagnostic method for automatic plaque characterisation in IVOCT.

For the development of the CNN model, patients and lesions from three different IVOCT studies[7-9] were included in the current *post hoc* analysis. Briefly, the studies included patients from five international centres located in Australia, USA, Japan, Spain and China, in both a retrospective[7,8] and a prospective fashion[9], with IVOCT performed for the evaluation of stable coronary lesions. Exclusion criteria were aorta-ostial lesions, bypass graft lesions, patients with moderate or severe valvular heart disease, acute coronary syndrome <72 hours attributed to the imaged vessel and chronic total occlusion in any other vessel. All IVOCT pullbacks were acquired with frequency-domain OCT systems of C7-XR™ or OPTIS™ (Abbott Vascular, Santa Clara, CA, USA) using a non-occlusive technique[10]. The institutional review boards of each individual centre approved the protocol of the studies, and all patients provided informed consent for enrolment in the institutional database for potential future investigations.

For the independent external validation, a different data set of 300 IVOCT images, provided by three international core labs, was used. Five patients were studied using the Lunawave® OFDI system (Terumo Corporation, Tokyo, Japan), while the rest were studied using the frequency-domain OCT OPTIS system.

DATA ANNOTATION

Ground truth was generated by labelling nine objects by experienced OCT analysts. The detailed annotation strategy is available in **Supplementary Appendix 1**.

DEEP CONVOLUTIONAL MODEL ARCHITECTURE: DESIGN AND TRAINING (Supplementary Appendix 2)

A U-shaped encoder–decoder architecture was designed, consisting of a contracting path for high-level feature extraction, an expansion path to produce full resolution segmentation, and vertical and horizontal feature bridges to preserve detailed spatial information **(Supplementary Figure 1)**. The model was fed with pseudo-3D input by stacking consecutive IVOCT cross-sections as separate colour channels to integrate the spatial information. A hybrid loss function of multi-class cross-entropy loss and focal Tversky loss was used to address the problem of class imbalance **(Supplementary Appendix 3)**. The detailed training strategy and ablation experiments to test the rationale of the CNN design are available in **Supplementary Appendix 4** and **Supplementary Appendix 5**.

MODEL DEVELOPMENT AND INTERNAL EVALUATION

The annotated pullbacks were randomly divided into training data set and testing data set, in a proportion of 9 to 1, strictly avoiding repetition of pullbacks in different data sets. The training set was used for the model development, of which 10% of the data set was separated for the hyper-parameter optimisation. After the model was fully developed, the testing data set was used for the internal evaluation of the model performance. The agreement between predictions and ground truth was evaluated by means of the Dice coefficient, calculated per category and then averaged over categories. The purity and completeness of positive predictions relative to the ground truths were reported as precision and recall, respectively.

Plaque burden (PB) was calculated as the area between the estimated internal elastic lamina (IEL) contour and lumen contour, divided by IEL area and multiplied by 100%[11]. The accuracy in segmentation of the IEL was appraised as the agreement in PB between the model prediction and the ground truth.

EXPERT CONSENSUS AND INDEPENDENT VALIDATION

The external validation of the model was performed on a different data set than the one used for the model development or internal evaluation. A total of 300 IVOCT images with different atherosclerotic plaque morphology and composition, acquired during clinical practice, were provided by three international core labs, each one providing 100 images (from 10 patients) with delimited regions **(Figure 1)**. Regions with poor signal penetration or insufficient quality, precluding adequate visualisation and analysis, were flagged with an arch-shaped annotation (**Figure 1**, region 3). Four experienced OCT readers (A. Maehera, Z. Ali, H. Jia, N. Holm) from the three core labs (CRF, New York, NY, USA; iMcorelab, 2nd Affiliated Hospital of Harbin Medical University, Harbin, China; Aarhus University Hospital, Skejby, Denmark) participated in the evaluation and labelled the regions with the corresponding tissue components, having access to the original images but blinded to each of the other analysts. If the evaluator considered that the delimited region had more than one tissue type, the region was then subdivided into multiple regions. According to the evaluation by the three core labs, each region was classified as "unanimity", "one core lab disagrees" or "all three core labs disagree". **Supplementary Figure 2** shows representative cases with unanimity and disagreement among core labs. Consensus was defined as agreement between ≥2 core labs, i.e., first two categories of the classification, and the consensus label for the region was accepted as reference for the independent validation.

The proposed deep convolutional model was integrated into the OctPlus software (Pulse Medical Imaging Technology, Shanghai, China) for real-time analysis of IVOCT pullbacks **(Figure 2)**, displaying plaques in both 2D cross-sectional and 3D views **(Figure 2A)**. The software quantified plaque area, arc degree and the different tissue proportions. An independent validation of the proposed CNN model within the software was performed, using the consensus regions as reference. Given that some plaque regions were flagged by arch-shaped annotations without complete boundaries, pixel-wise evaluation by Dice coefficient was not applicable for the external validation. Model performance was reported by the means of accuracy. Correct plaque characterisation was defined as a >80% overlapping portion between the prediction of the software and the core lab consensus in area or arc circumference. Sensitivity analysis on the impact of the overlapping portion on the diagnostic accuracy was performed by changing the overlapping portion from 80% to 90%.

STATISTICAL ANALYSIS

A normality test was performed using the Shapiro-Wilk test. Continuous variables are presented as mean±SD or median (interquartile range), as appropriate, whereas categorical variables are presented as counts and percentages. Correlation between ground truth and model predictions was evaluated using Pearson or Spearman correlation tests, as appropriate. Agreement between groups for continuous variables was assessed by means of Bland-Altman analysis and intraclass correlation coefficient for the absolute agreement (ICCa), whilst the kappa coefficient was used for categorical variables. A two-sided p-value ≤0.05 was considered statistically significant. A confidence level of 95% (95% CI) was used to estimate the plausible range of values. Statistical analysis was performed using SPSS, Version 23.0 (IBM Corp., Armonk, NY, USA).

Results

STUDY POPULATION CHARACTERISTICS

A total of 509 OCT pullbacks from 391 patients were analysed, resulting in 10,517 and 1,156 cross-sections for the training and testing data sets, respectively **(Figure 3)**. Patient baseline clinical characteristics are presented in **Table 1**.

INTERNAL EVALUATION

The model performance on the testing data set is summarised in **Table 2**. The model performed the best segmentation on fibrous plaque (Dice=0.906), followed by calcific (Dice=0.848) and lipidic plaque (Dice=0.772). For the segmentation of markers

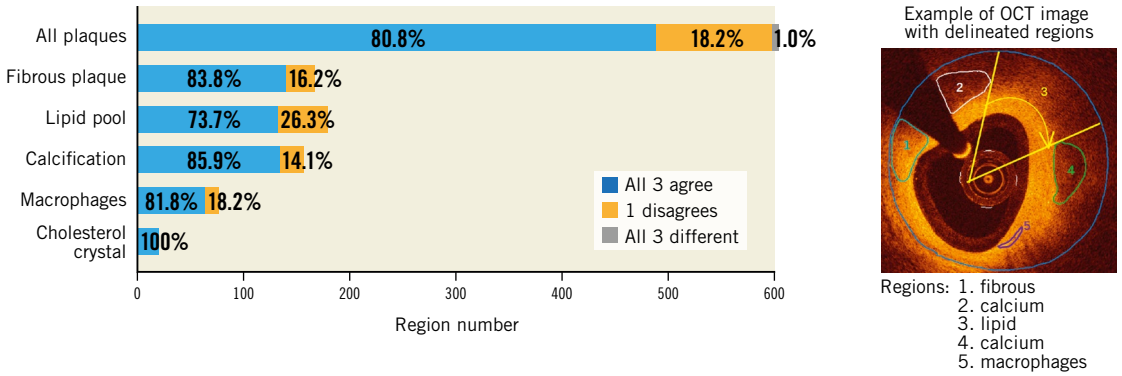


Figure 1. *Inter-core lab agreement and variability on plaque characterisation. Example of an IVOCT image with delimited regions (marked by numbers) and results of plaque characterisation, stratified by the different tissue components, as determined by the three core labs. There is good agreement for the majority of calls, since all three or at least two core labs agreed on the diagnosis.*

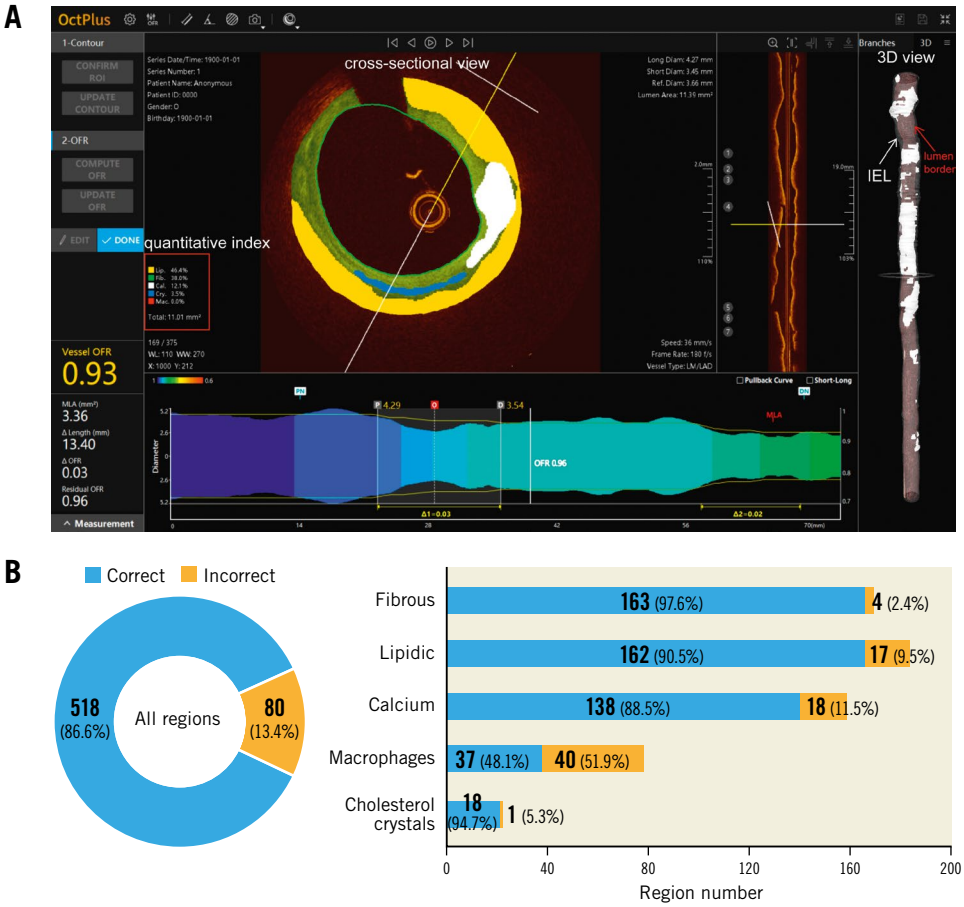


Figure 2. *The proposed AI model was integrated into the OctPlus software and externally validated. A) Screen shot of plaque quantitative assessments by the software. 3D mapping of calcifications between IEL (white arrow) and lumen contour (red arrow) is shown on the right side. B) Diagnostic accuracy of the software, stratified by the different plaque components, taking the inter-core lab consensus as standard reference. AI: artificial intelligence; IEL: internal elastic lamina*

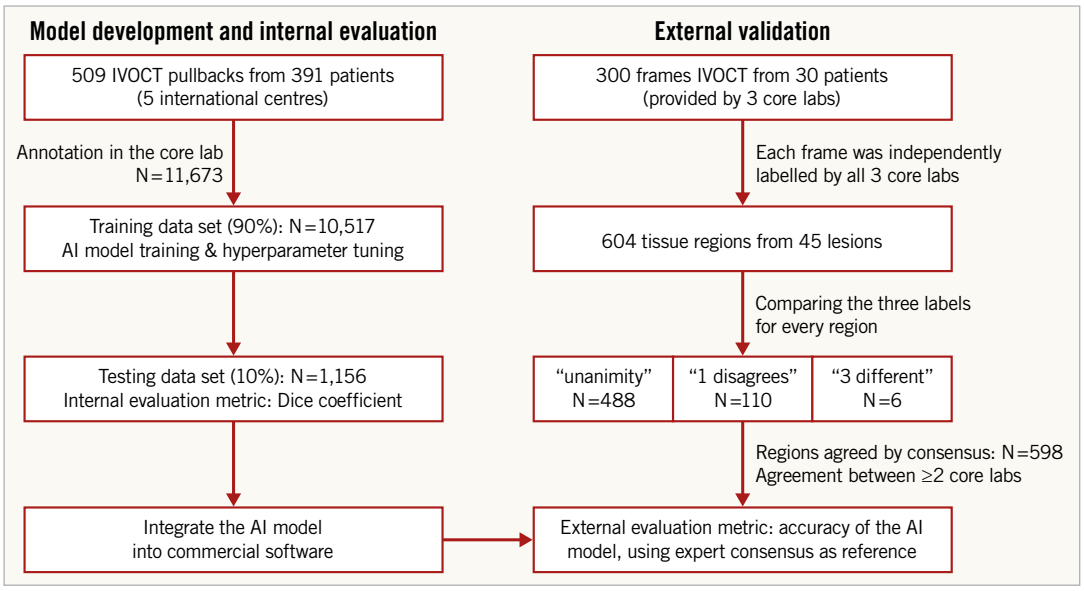


Figure 3. *Study flow chart. The utility of the different data sets for internal evaluation (left column) and external validation (right column) is indicated.*

Table 1. Baseline clinical characteristics.

Patients (N=391)	
Age, years	66.4±10.7
Male	298 (76.2%)
BMI, kg/m^2	25.9±5.0
Diabetes mellitus	158 (40.4%)
Hypertension	311 (79.5%)
Hyperlipidaemia	265 (67.8%)
Current smoker	100 (25.6%)
Family history of CAD	78 (20.5%)
Previous PCI	220 (56.3%)
Previous CABG	6 (1.5%)
Previous MI	156 (39.9%)
Clinical presentation	
Silent ischaemia	128 (32.7%)
Stable angina	118 (30.2%)
Unstable angina	67 (17.1%)
NSTEMI	14 (3.6%)
Other	64 (16.4%)

Data are presented as mean±SD, median (Q1-Q2), or n (%), as appropriate. BMI: body mass index; CABG: coronary artery bypass graft; CAD: coronary artery disease; MI: myocardial infarction; NSTEMI: non-ST-elevation myocardial infarction; PCI: percutaneous coronary intervention

Table 2. Segmentation performance of the proposed model on the testing data set.

		Precision	Recall	Dice
Plaques	Fibrous	0.932	0.881	0.906
	Lipidic	0.739	0.807	0.772
	Calcific	0.811	0.888	0.848
Markers of complicated plaque	Cholesterol crystals	0.543	0.508	0.525
	Macrophage	0.429	0.568	0.489
	Microvessel	0.598	0.604	0.601
Other structures	Guidewire artefact	0.870	0.920	0.894
	Side branch	0.815	0.890	0.851
	IEL	0.989	0.989	0.989
Mean		0.747	0.784	0.764

IEL: internal elastic lamina

of inflammation/complicated plaque, the model performed best on microvessels (Dice=0.601), followed by cholesterol crystals (Dice=0.525) and macrophages (Dice=0.489). Segmentation of non-tissue structures (i.e., guidewire and side branches) also achieved high performance **(Table 2)**. Among quantitative parameters, the PB assessed by the model correlated very well with the ground truth (R2=0.98, p<0.001) **(Figure 4)**, with a mean difference of 0.35±2.2% and an ICCa of 0.99 (95% CI: 0.98-0.99).

Figure 5 shows examples of the model predictions in different challenging scenarios. The model had excellent performance on cases with insufficient blood clearance (row A), complex calcification (row B) or large lipidic burden (row C), while displaying certain ability in segmenting different inflammatory markers, like cholesterol crystals, macrophages or microvessels (row D).

INTER-CORE LAB AGREEMENT AND VARIABILITY IN PLAQUE CHARACTERISATION

For the external validation, a total of 45 lesions, including 10 long lesions (>28 mm) and 6 diffuse lesions from the 30 IVOCT examinations were analysed, resulting in 604 plaque regions labelled by the core labs **(Figure 3)**. The median lesion length and minimal lumen area of the lesions were 20.40 (10.90-33.05) mm and 2.00 (1.36-3.00) mm^2, respectively. Consensus on plaque characterisation was reached in 598 (99% [95% CI: 97.8-99.6%]) regions **(Figure 1)**; unanimity among core labs was observed in 488 (81% [95% CI: 77.5-83.7%]) regions, and agreement between two core labs in 110 (18% [95% CI: 15.3-21.5%]) regions. Unanimity

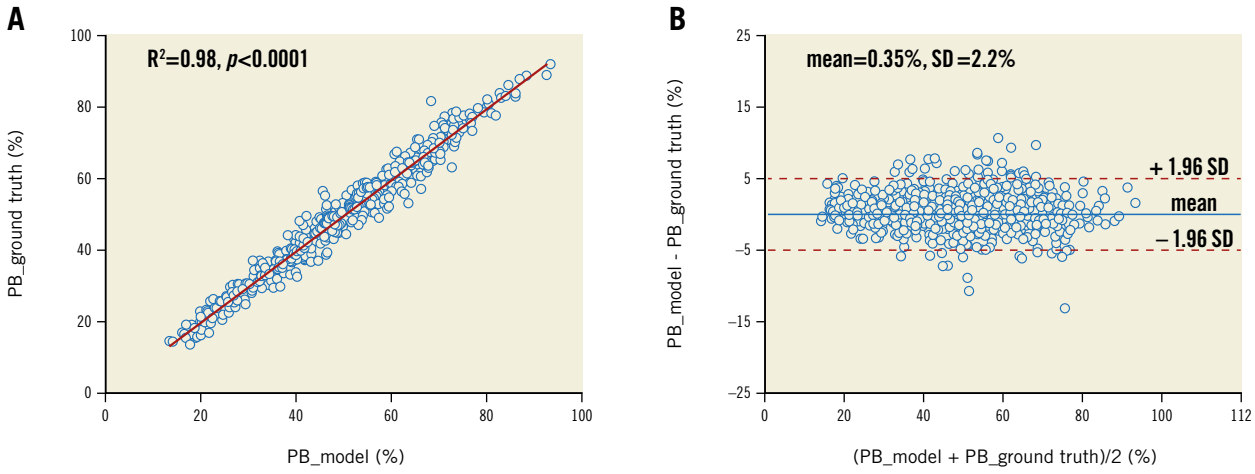


Figure 4. *Agreement of plaque burden (PB) between AI predictions and manual annotations. A) Correlation in PB between model prediction and ground truth. B) Bland-Altman analysis. Middle blue line: mean difference; red dotted lines: mean±1.96 SD. PB: plaque burden*

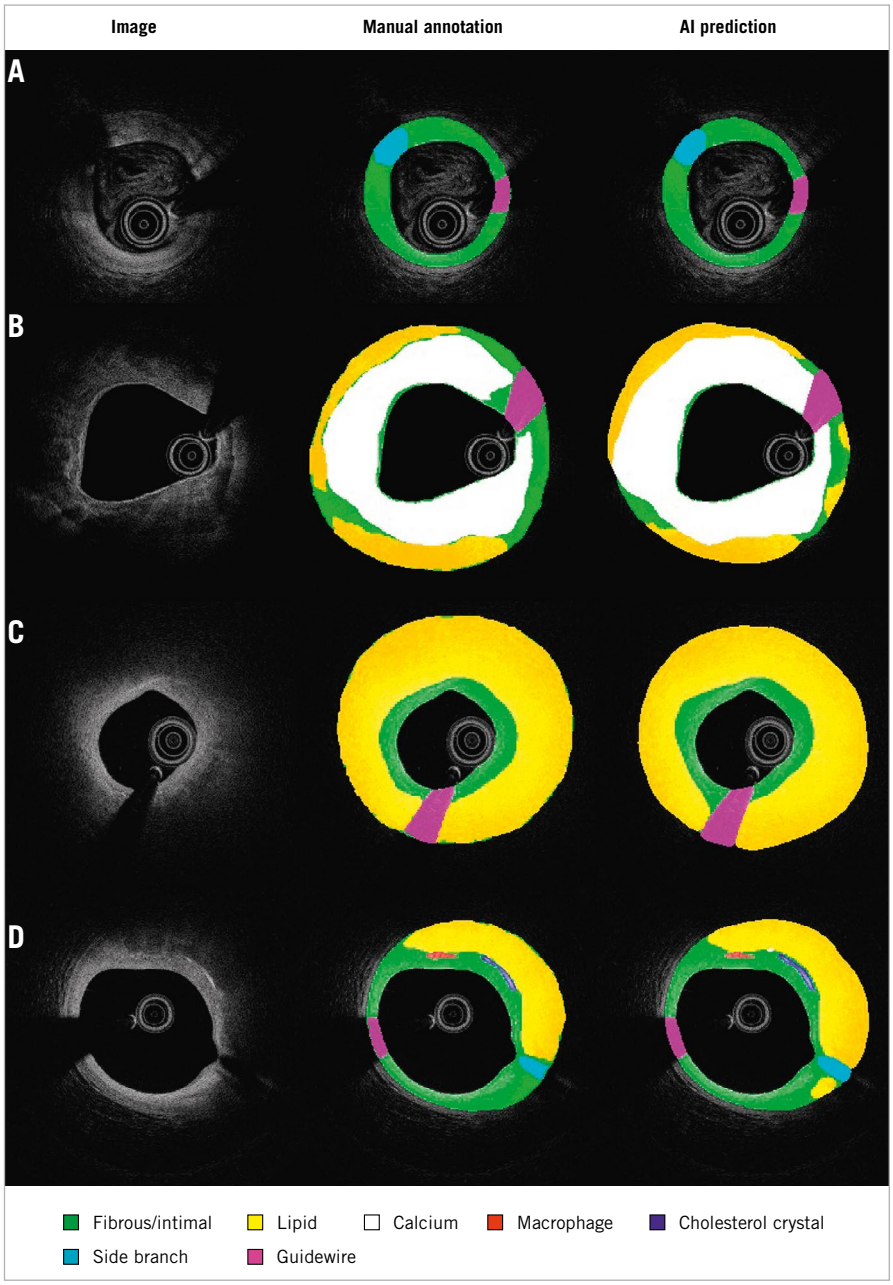


Figure 5. *Segmentation in different challenging situations. Images with suboptimal quality (A), heavy calcification (B), plaques with large lipidic pool (more than two quadrants) (C), inflammatory markers (D).*

among core labs was most frequently observed for cholesterol crystals (100% [95% CI: 82.4-100%]), followed by calcific plaque (85.9% [95% CI: 79.5-90.6%]), fibrous plaque (83.8% [95% CI: 77.5-88.7%]), macrophages (81.8% [95% CI: 71.6-88.9%]), and lipidic plaque (73.7% [95% CI: 66.8-79.7%]). The data on agreement between individual core labs with consensus are available in **Supplementary Appendix 6** and **Supplementary Figure 3**.

EXTERNAL VALIDATION

In the external validation, the software correctly segmented and characterised 518 out of 598 regions agreed by consensus, corresponding to an overall diagnostic accuracy of 86.6% (95% CI: 83.7-89.1%). The software performed the best in fibrous plaque (accuracy 97.6% [95% CI: 93.4-99.3%]), followed by lipidic plaque (90.5% [95% CI: 85.2-94.1%]) and calcifications (88.5% [95% CI: 82.4-92.7%]). Cholesterol crystals were also well characterised (accuracy 94.7% [95% CI: 73.5-100%]), but the performance for macrophages was suboptimal (48.1% [95% CI: 37.3-59.0%]) **(Figure 2B)**. The overall diagnostic accuracy was numerically higher in unanimous regions where all three core labs agreed than in those regions where only two core labs agreed: 89.7% (95% CI: 86.7-92.2%) versus 72.7% (95% CI: 63.7-80.2%), p<0.001. When the threshold of overlapping area to define a correct characterisation between the software prediction

and the consensus increased from 80% to 90%, diagnostic accuracy decreased slightly but remained high, with 96.4% (95% CI: 92.2-98.5%) for fibrous plaque, 87.2% (95% CI: 81.4-91.3%) for lipidic plaque, and 85.9% (95% CI: 79.5-90.6%) for calcifications.

The diagnostic accuracy was similar in both OCT systems for basic plaque components, 91.9% (95% CI: 88.9-94.2%) in Abbott versus 94.2% (95% CI: 85.6-98.2%) in Terumo (p=0.63), and for all tissue regions 87.5% (95% CI: 84.3-90.1%) in Abbott versus 81.8% (95% CI: 72.4-88.6%) in Terumo (p=0.17).

MODEL RATIONALE AND ANALYSIS TIME

Systematic ablation experiments using the testing data set verified the rationale in the design of the deep convolutional model, including pseudo-3D input, vertical feature bridges, and hybrid loss **(Supplementary Table 1)**.

The median time required for the CNN model to analyse an image pullback with 271 (263-375) cross-sections of 704×704 pixel size was 21.4 (18.6-25.0) seconds, corresponding to an average speed of 0.07±0.01 seconds per cross-section using a laptop equipped with AMD Ryzen 7 and Geforce RTX 2060 graphic card.

Discussion

The following points summarise the key findings of the present study. 1) An AI model based on a deep convolutional neural network for automatic IVOCT plaque characterisation was developed and validated, proving fast computational speed and excellent performance in a real-world series of images. 2) Consensus on coronary tissue characterisation using OCT could be achieved in most of the plaque regions among the three core labs, 81% with unanimity among three core labs and 18% with agreement between two core labs. 3) The AI model performed the best in fibrous plaque, followed by lipidic plaque and calcifications, with diagnostic accuracy of 97.6%, 90.5% and 88.5%, respectively. 4) The plaque burden automatically assessed by the AI model correlated well with the core lab analysis (R2=0.98, p<0.001).

Several studies have previously attempted the automatic characterisation of atherosclerotic plaque on IVOCT using AI[4,5]. However, these studies developed their models on relatively small data sets with limited diversity, thus entailing problems of generalisability for clinical use. Furthermore, most of these models were validated pursuant to internal analysis within the team, lacking an independent external assessment based on expert consensus. Additionally, the *a priori* knowledge of spatial continuity along adjacent frames had not been fully exploited and only a few tissue components had been qualitatively characterised hitherto, with limited quantification. In addition to intravascular imaging, the applications of AI in the field of plaque characterisation using cardiac computed tomography (CT) are also expanding. Multiple approaches including machine learning and CNN models have been proposed for automatic calcium detection and scoring using CT images[12,13].

From a technical point of view, this novel AI model is unique in many aspects, including the integration of spatial information from contiguous cross-sections to enhance the diagnostic accuracy by means of pseudo-3D input, the incorporation of multi-scale feature forward bridges in both horizontal and vertical directions for better fusion of features, and the use of a hybrid loss function to address challenges in segmentation of inflammatory markers. The rationale for the design of this model was verified on the testing data set **(Supplementary Table 1)**. It is important to note that the average analysis speed for the model is 0.07±0.01 seconds per cross-section while it took the analysts several minutes to annotate one frame in the core lab. It would be more time-demanding for images with complex plaque compositions or suboptimal quality since the specialists need to evaluate cross-sections from several adjacent imaging slices while the model rapidly integrated the information of spatial continuity across frames.

The development of the AI model was focused on clinical applications. In this regard it is important to highlight that the model was trained and validated with a large volume of IVOCT images, encompassing a range of image quality, plaque composition and lesion complexity in a representative population with ischaemic heart disease. This characteristic of the training data set is crucial to guarantee the generalisability of the CNN model. Furthermore, the external validity of the study was retested against a high-quality reference standard, the consensus of three leading international core labs, with excellent diagnostic concordance achieved for most plaque types. Although the data set for external validation was modest in size, it was of different complexity and sufficient to assess the external validity of our findings.

The AI model performed very well in the segmentation of basic plaque components in both internal evaluation and external validation, while the diagnostic accuracy was only modest for markers of high risk and complex plaques. In both categories, the agreement with the expert consensus was better in structures with low versus high attenuation, i.e., the model was more accurate in fibrous than in lipidic tissue, and more accurate in cholesterol crystals than in macrophages. These findings will require a specific appraisal in future studies but might be partially explained by the agreement between the experts in the different categories. The experts tended to disagree the most in structures with high attenuation (lipidic pools, macrophages) or in regions lying very abluminally, where the signal was poor, the quality of the image was lower and therefore the subjective interpretation of the analyst played a greater role. Indeed, the AI model performed significantly better in regions with unanimity than in regions without unanimity (89.7% vs 72.7%, p<0.001).

Of note, the accuracy of the AI model in identifying lipidic plaques (90.5%) was slightly higher than in calcified plaques (88.5%) in the external validation. This was because some lipidic plaques with poor signal penetration, precluding adequate visualisation of boundaries, were flagged with an arch-shaped annotation, which reduced the difficulty for the AI model to meet the standards compared to the calcium plaques delineated with complete boundaries. The accuracy in identifying cholesterol crystals was excellent in the external validation while segmentation performance was modest in the internal testing data set. This might be explained by the small area

occupied by cholesterol crystals; a small change in the segmentation border would impact significantly on the Dice index, while preserving good agreement in tissue characterisation. Nevertheless, the model might require further fine-tuning to improve its performance in categories such as macrophages.

To the best of our knowledge, this is the first IVOCT study to report the reproducibility of outlining the IEL. This was a challenging task for IVOCT, because the media often lies beyond IVOCT penetration in thick plaques or is hidden behind lipid-rich pools, causing intense attenuation of the signal[1]. Nevertheless, recent studies have proven that IVOCT analysts can identify the external elastic lamina (EEL) at >180º of the circumference in 95% of the cross-sections in a core lab setting[14]. Thus, considering the circular geometry of the arterial structures in the cross-section and the information from continuous frames, the invisible part of IEL and EEL can be reliably extrapolated. As shown in **Figure 6**, the predicted IEL by our proposed model shows good concordance with an adjacent reference cross-section, even in lesions with a large lipidic pool. This is a key step forward for the quantitative assessment of atherosclerosis by means of IVOCT, with potential prognostic implications, as plaque burden is an independent predictor of future events in non-culprit coronary lesions[15].

Plaque composition also has prognostic value, as large lipidic burden and thin-cap fibroatheroma (TCFA) are associated with higher incidence of periprocedural myocardial infarction[16]. The current AI model provides comprehensive *in vivo* analysis of atherosclerotic plaque composition and morphological features of plaque progression or instability in a fully automated fashion, thus

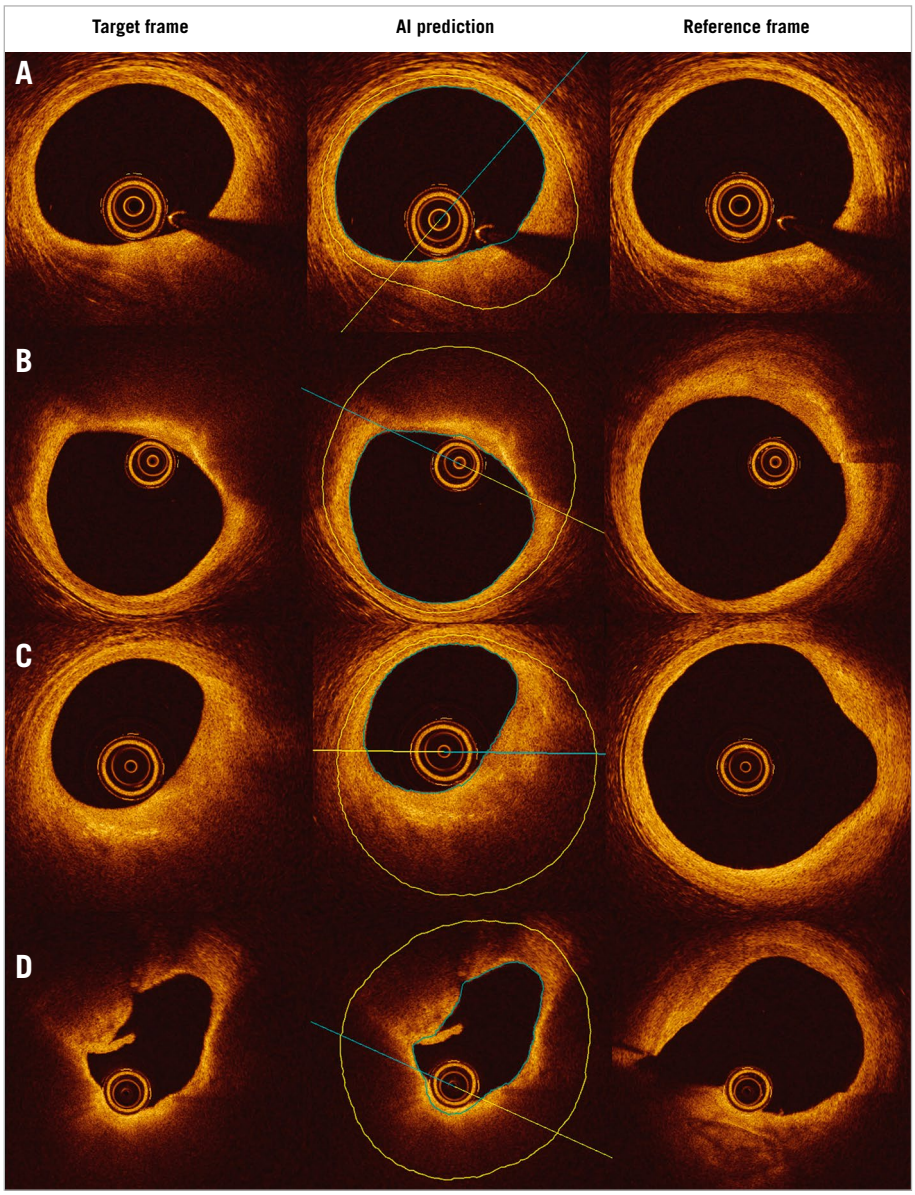


Figure 6. *Predicted IEL by AI model shows good concordance with reference cross-section in plaques with small (A), medium (B, C) and large lipidic pool (D). IEL: internal elastic lamina*

providing interventional cardiologists, irrespective of their imaging expertise, with the same level of proficiency as top imaging experts, while sparing time and weariness. The CLIMA trial has recently shown that the identification of macrophages, together with other vulnerable plaque features on OCT, has prognostic value in predicting the population at high risk of acute events[17]. The AI model is able to identify inflammatory markers automatically along the entire IVOCT pullback rather than in a small region of interest, thus providing a more comprehensive assessment that might potentially result in better risk stratification than conventional image interpretation.

Limitations

This was a study to validate an AI model of automatic segmentation and characterisation of plaque composition, based on the consensus of IVOCT experts. It does not intend to provide a histological validation of IVOCT for plaque characterisation, which has already been achieved[18]. Of note, histological validation in humans can only be performed by autopsy in cadavers, in patients who died from acute cardiac events. Obtaining histological data from stable coronary lesions in patients dying from non-cardiac causes has proven problematic.

Considering the potential clinical and research applications of the model, the study focused on stable atherosclerotic plaques, excluding lesions with signs of instability (thrombus, plaque rupture, dissection or haematoma). In addition, the presence of metallic stents generates dark trailing shadows on OCT images, jeopardising the interpretation of plaque composition. Thus, cross-sectional images at the stented segment were excluded from the present analysis. The results of the model in unstable plaques and in stented vessel segments should be interpreted with caution and would require further investigation in specific studies.

Conclusions

A novel AI framework for automatic plaque characterisation in IVOCT was developed, providing excellent diagnostic accuracy in both internal and external validation. This model might reduce subjectivity in image interpretation and facilitate IVOCT quantification of plaque composition, with potential applications in research and IVOCT-guided PCI.

Impact on daily practice

An AI-based model was developed and validated for automatic plaque characterisation on IVOCT. This substantially improved the objectivity and reproducibility of IVOCT quantification. The AI model enables comprehensive plaque characterisation and identification of inflammatory markers, creating an interesting perspective for future studies on plaque progression and risk stratification. The model has the potential to assist IVOCT-guided PCI by tailoring the intervention according to plaque composition and by using internal elastic lamina as reference for stent sizing.

Funding

This work was supported by the National Key Research and Development Program of China, Natural Science Foundation of China (Grant Number 81871460, 82020108015, 81827806 and 81671763), and by a Science Foundation Ireland Research Professorship Award (RSF 1413).

Conflict of interest statement

Z. Ali reports institutional grants from NIH/NHLBI, Abbott Vascular, and Cardiovascular Systems Inc., personal fees from Amgen, AstraZeneca, and Boston Scientific, and personal equity in Shockwave Medical, outside the submitted work. G. Mintz has received honoraria from Boston Scientific, Philips/Volcano, Medtronic, and Terumo. W. Wijns reports an institutional research grant and honoraria from MicroPort; he is a co-founder of Argonauts, an innovation facilitator. N. Holm has received institutional research grants from Boston Scientific, Biosensors, Abbott, Reva Medical and Medis medical imaging, and speaker fees from Abbott, Terumo, Medis medical imaging and Reva Medical. S. Tu has received research support from Pulse Medical Imaging Technology. The other authors have no conflicts of interest to declare.

References

1. Gutiérrez-Chico JL, Alegría-Barrero E, Teijeiro-Mestre R, Chan PH, Tsujioka H, de Silva R, Viceconte N, Lindsay A, Patterson T, Foin N, Akasaka T, di Mario C. Optical coherence tomography: from research to practice. *Eur Heart J Cardiovasc Imaging.* 2012;13:370-84.

2. Gutiérrez-Chico JL, Cortés C, Schincariol M, Limon U, Yalcinli M, Durán-Cortés MA, Jaguszewski M. Implantation of bioresorbable scaffolds under guidance of optical coherence tomography: feasibility and pilot clinical results of a systematic protocol. *Cardiol J.* 2018;25:443-58.

3. Wijns W, Shite J, Jones MR, Lee SWL, Price MJ, Fabbiocchi F, Barbato E, Akasaka T, Bezerra H, Holmes D. Optical coherence tomography imaging during percutaneous coronary intervention impacts physician decision-making: ILUMIEN I study. *Eur Heart J.* 2015;36:3346-55.

4. Gessert N, Lutz M, Heyder M, Latus S, Leistner DM, Abdelwahed YS, Schlaefer A. Automatic plaque detection in IVOCT pullbacks using convolutional neural networks. *IEEE T Med Imaging.* 2018;38:426-34.

5. Abdolmanafi A, Duong L, Dahdah N, Cheriet F. Deep feature learning for automatic tissue classification of coronary artery using optical coherence tomography. *Biomed Opt Express.* 2017;8:1203-20.

6. Li L, Jia T. Optical Coherence Tomography Vulnerable Plaque Segmentation Based on Deep Residual U-Net. *Rev Cardiovasc Med.* 2019;20:171-7.

7. Yu W, Huang J, Jia D, Chen S, Raffel OC, Ding D, Tian F, Kan J, Zhang S, Yan F, Chen Y, Bezerra HG, Wijns W, Tu S. Diagnostic accuracy of intracoronary optical coherence tomography-derived fractional flow reserve for assessment of coronary stenosis severity. *EuroIntervention.* 2019;15:189-97.

8. Huang J, Emori H, Ding D, Kubo T, Yu W, Huang P, Zhang S, Gutiérrez-Chico J, Akasaka T, Wijns W, Tu S. Diagnostic performance of intracoronary optical coherence tomography-based and angiography-based fractional flow reserve for evaluation of coronary stenosis. *EuroIntervention.* 2020;16:568-76.

9. Gutiérrez-Chico JL, Chen Y, Yu W, Ding D, Huang J, Huang P, Jing J, Chu M, Wu P, Tian F, Xu B, Tu S. Diagnostic accuracy and reproducibility of optical flow ratio for functional evaluation of coronary stenosis in a prospective series. *Cardiol J.* 2020; 27:350-61.

10. Prati F, Cera M, Ramazzotti V, Imola F, Giudice R, Albertucci M. Safety and feasibility of a new non-occlusive technique for facilitated intracoronary optical coherence tomography (OCT) acquisition in various clinical and anatomical scenarios. *EuroIntervention.* 2007;3:365-70.

11. Sangiorgi G, Rumberger JA, Severson A, Edwards WD, Gregoire J, Fitzpatrick LA, Schwartz RS. Arterial calcification and not lumen stenosis is highly correlated with atherosclerotic plaque burden in humans: a histologic study of 723 coronary artery segments using nondecalcifying methodology. *J Am Coll Cardiol.* 1998;31:126-33.

12. Wolterink JM, Leiner T, Takx RA, Viergever MA, Išgum I. Automatic Coronary Calcium Scoring in Non-Contrast-Enhanced ECG-Triggered Cardiac CT With Ambiguity Detection. *IEEE Trans Med Imaging.* 2015;34:1867-78.

13. Kolossváry M, De Cecco CN, Feuchtner G, Maurovich-Horvat P. Advanced atherosclerosis imaging by CT: Radiomics, machine learning and deep learning. *Cardiovasc Comput Tomogr*. 2019;13:274-80.

14. Ali ZA, Maehara A, Généreux P, Shlofmitz RA, Fabbiocchi F, Nazif TM, Guagliumi G, Meraj PM, Alfonso F, Samady H, Akasaka T, Carlson EB, Leesar MA, Matsumura M, Ozan MO, Mintz GS, Ben-Yehuda O, Stone GW; ILUMIEN III: OPTIMIZE PCI Investigators. Optical coherence tomography compared with intravascular ultrasound and with angiography to guide coronary stent implantation (ILUMIEN III: OPTIMIZE PCI): a randomised controlled trial. *Lancet.* 2016;388:2618-28.

15. Stone GW, Maehara A, Lansky AJ, De Bruyne B, Cristea E, Mintz GS, Mehran R, McPherson J, Farhat N, Marso SP, Parise H, Templin B, White R, Zhang Z, Serruys PW; PROSPECT Investigators. A prospective natural-history study of coronary atherosclerosis. *N Engl J Med.* 2011;364:226-35.

16. Kini AS, Motoyama S, Vengrenyuk Y, Feig JE, Pena J, Baber U, Bhat AM, Moreno P, Kovacic JC, Narula J, Sharma SK. Multimodality Intravascular Imaging to Predict Periprocedural Myocardial Infarction During Percutaneous Coronary Intervention. *JACC Cardiovasc Interv.* 2015;8:937-45.

17. Prati F, Romagnoli E, Gatto L, La Manna A, Burzotta F, Ozaki Y, Marco V, Boi A, Fineschi M, Fabbiocchi F, Taglieri N, Niccoli G, Trani C, Versaci F, Calligaris G, Ruscica G, Di Giorgio A, Vergallo R, Albertucci M, Biondi-Zoccai G, Tamburino C, Crea F, Alfonso F, Arbustini E. Relationship between coronary plaque morphology of the left anterior descending artery and 12 months clinical outcome: the CLIMA study. *Eur Heart J.* 2020;41:383-91.

18. Yabushita H, Bouma BE, Houser SL, Aretz HT, Jang IK, Schlendorf KH, Kauffman CR, Shishkov M, Kang DH, Halpern EF, Tearney GJ.. Characterization of human atherosclerosis by optical coherence tomography. *Circulation.* 2002;106:1640-5.

Supplementary data

Supplementary Appendix 1. Data annotation strategy.

Supplementary Appendix 2. Deep convolutional model architecture.

Supplementary Appendix 3. Proposed hybrid loss.

Supplementary Appendix 4. Training of the deep convolutional model.

Supplementary Appendix 5. Ablation experiments.

Supplementary Appendix 6. Inter-core lab agreement and variability in plaque characterisation.

Supplementary Figure 1. Schematic diagram of the model architecture.

Supplementary Figure 2. Cases of plaque characterisation by core labs in the external validation data set.

Supplementary Figure 3. Consensus on plaque tissue types and agreement by the three core labs.

Supplementary Table 1. Ablation study results of all regions on the testing data set.

The supplementary data are published online at: https://eurointervention.pcronline.com/doi/10.4244/EIJ-D-20-01355

Supplementary data

Supplementary Appendix 1. Data annotation strategy

Ground truth was generated by labelling the plaque components in cross-sections at 2 mm longitudinal intervals in the IVOCT pullbacks. Annotation was manually performed by experienced OCT analysts (M. Chu, X. Zeng and M. Zeng), trained and regularly audited in a core laboratory (CardHemo, Med-X Research Institute, Shanghai Jiao Tong University, Shanghai, China). Lumen contours and internal elastic lamina (IEL) were outlined, and tissues in between were classified as (a) non-tissue parts: guidewire artefacts or side branches; (b) plaque components: fibrous tissue, lipidic pool or calcium; (c) markers of plaque instability and inflammation: macrophages, cholesterol crystals or microvessels. When several plaque components were observed in the same cross-section, each component was delineated separately. Cross-sections with plaque rupture, thrombus, stent struts, or insufficient quality were excluded from the analysis. Sectors with normal intima were annotated as fibrous tissue. All annotations were verified by an OCT specialist (L. He), who has clinically interpreted more than 2,000 OCT pullbacks under the supervision of a senior cardiologist (H. Jia). The annotated pullbacks were randomly divided into training set and testing set, in a proportion of 9 to 1, strictly avoiding repetition of pullbacks in different data sets.

Supplementary Appendix 2. Deep convolutional model architectures

The model followed a U-shaped encoder–decoder architecture composed of contracting path, expansion path, vertical and horizontal feature bridges **(Supplementary Figure 1)**. The encoder path consisted of repeated applications of a convolutional module, comprising two 3x3 convolution layers for feature extraction, each followed by a non-linear activation (ReLu). Adjacent convolutional modules were connecting by means of a 2x2 max pooling operation for down-sampling and a batch normalisation layer. In the decoder path, the feature maps were up-sampled through a symmetric structure with the contracting path by repetition of the convolutional module and up-sampling operations to produce a full-resolution segmentation. Compared to the standard U-net model, the proposed model had superior depth of network and less feature maps at the corresponding module by a factor of ½, resulting in a deeper and thinner model. In addition to the horizontal shortcuts between contracting and expansion layers as in U-net, the encoding path also incorporated vertical feature bridges that forward high-resolution local features of the input to deeper convolutional modules at multi-scales. At the final layer, a 1x1 convolution and softmax function was used to generate multi-class probability maps.

Of note, the model feeds a 2D CNN model with a stack of consecutive IVOCT cross-sections as pseudo-3D input, to integrate the spatial information in the longitudinal direction and produce the segmentation of the central frame. Furthermore, the vertical feature bridges forward the local spatial information of pseudo-3D images to deeper layers at multi-scales, thus refining the segmentation results by means of recovering the spatial details lost in the aggressive down-sampling at the encoding path.

Supplementary Appendix 3. Proposed hybrid loss

The segmentation of macrophages, cholesterol crystals and microvessels was technically difficult because they were small in size and infrequent, causing severe class imbalance when the model also needs to segment other plaque components. Objects of rare classes have negligible weight in the loss function, resulting in biased models with high false negativity. Special loss functions have been proposed to circumvent this problem. Focal loss forces the model to learn the poorly classified pixels better, focusing on challenging, misclassified examples. Focal loss was initially introduced for detection tasks, rather than for segmentation. Dice loss (the negative of Dice coefficient) was initially used as loss function for volumetric segmentation of medical images. It can partially address the problem of data imbalance by turning pixel-wise loss composition (every pixel contributes to the loss) into class-level distribution distance (loss is calculated at a category level). Tversky loss is a generalisation of Dice loss, able to tune trade-offs between false positives and false negatives, but segmentations based on class-level loss alone may result in rough borders. In order to address the class imbalance problem and stabilise the training process, a combination of multi-class cross-entropy loss (Equation 1) and focal Tversky loss (Equation 2) was proposed in the current study. Similar to focal loss, focal Tversky loss was derived from Tversky score and underweights the contribution of easy categories to the loss during the training, focusing on challenging categories.

$$Loss_{CE} = -\frac{1}{N}\sum_{c=1}^{C}\sum_{n=1}^{N} g_{nc} \log p_{nc} \quad \text{(Equation 1)}$$

$$Loss_{FT} = \sum_{c=1}^{C}\left(1 - \frac{TP_c}{TP_c + \alpha \cdot FN_c + \beta \cdot FP_c}\right)^{\gamma} \quad \text{(Equation 2)}$$

$$Loss = \lambda Loss_{CE} + Loss_{FT} \quad \text{(Equation 3)}$$

In the equations, C represents the total categories, N the image pixel numbers, p_{nc} and g_{nc} are the predicted possibility and ground truth value for pixel n of category c; TP_c, FN_c and FP_c are true positives, false negatives and false positives of predictions for category c; α and β are the trade-offs of penalties for false negatives and false positives. Hyper-parameter γ focuses the model on poorly performed

categories and was set to 0.75, while λ is a magnitude balance between focal Tversky loss and cross-entropy loss, which was set as 10 in this study.

Supplementary Appendix 4. Training of the deep convolutional model

OCT images were converted to greyscale prior to the training and validation process. The annotated cross-section and the immediately adjacent ones, together with their lumen images, were stacked and fed into the model as input. Tailored data augmentation was performed on-the-fly, involving random rotation, scale, panning, vertical and horizontal flip. A total of 250 epochs were performed using the Adam optimiser.

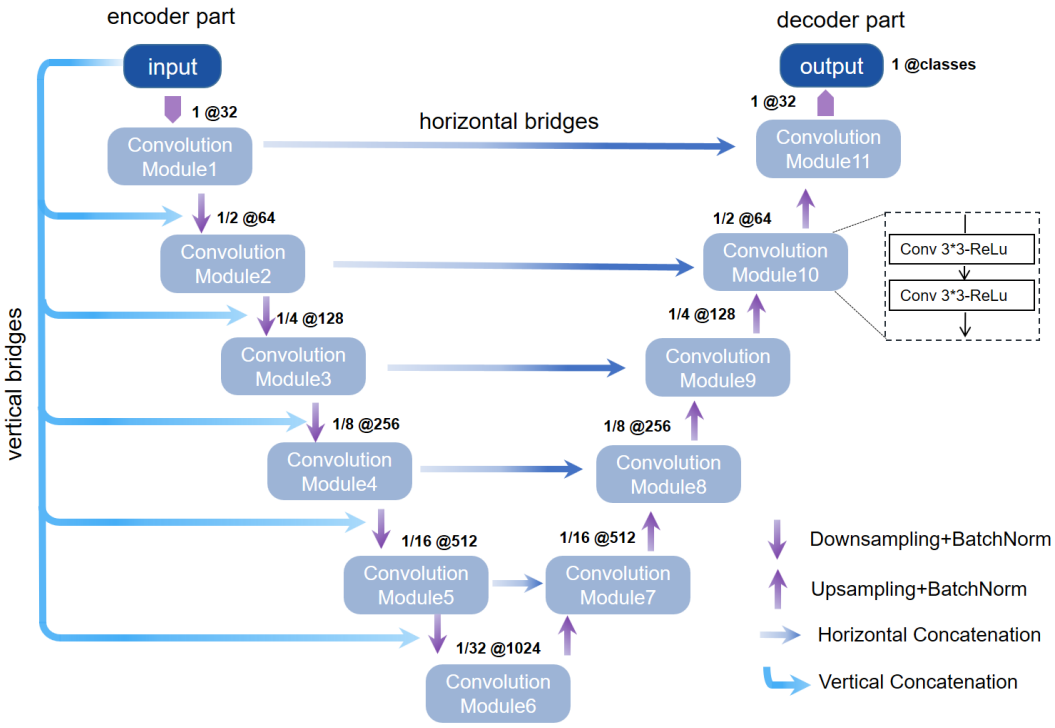
Supplementary Appendix 5. Ablation experiments

Systematic ablation experiments were carried out on the testing data set to prove the rationale behind the design of the deep convolutional model, comparing: 1) pseudo-3D input with different numbers of stacked frames versus single 2D cross-sectional input; 2) models with versus without vertical feature bridges; 3) the proposed hybrid loss (Equation 3) versus multi-class cross-entropy loss.

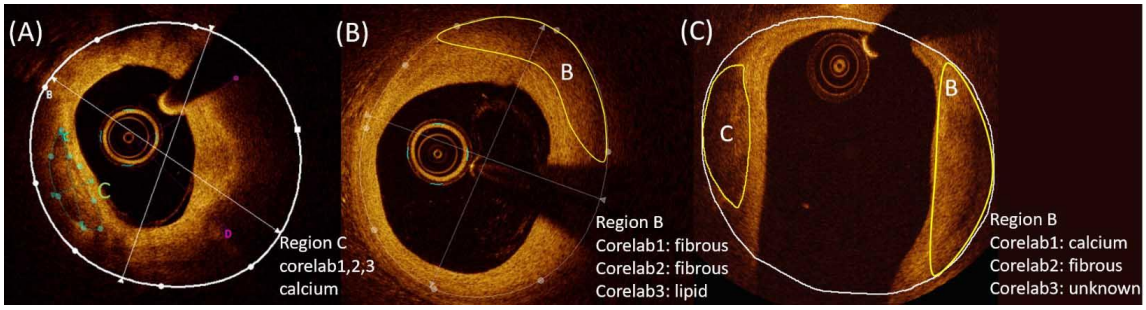
The results of the ablation study are summarised in **Supplementary Table 1**. Pseudo-3D input with consecutive images outperformed single cross-sectional 2D input. The evaluation metrics gradually improved in parallel to the number of stacked cross-sections up to a point (7-9 cross-sections) and then decreased. The addition of vertical feature bridges further improved the segmentation precision to 0.817 but showed relatively low recall value of 0.650. The adoption of the hybrid loss addressed this issue by tuning trade-offs between false negatives and false positives, achieving relatively balanced precision and recall.

Supplementary Appendix 6. Inter-core lab agreement and variability in plaque characterisation

Supplementary Figure 2 shows representative cases with unanimity and disagreement among core labs. **Supplementary Figure 3** shows the distributions of the different plaque components in the different core labs. Kappa coefficients for the different core labs, taking the consensus as reference, were 0.89 (95% CI: 0.86-0.92, $p<0.001$), 0.96 (95% CI: 0.94-0.98, $p<0.001$), and 0.88 (95% CI: 0.85-0.91, $p<0.001$).



Supplementary Figure 1. Schematic diagram of the model architecture.
The proposed model architecture consists of encoder path, decoder path, and vertical and horizontal feature bridges. Numbers next to the convolution module represent the size of feature maps, where the number after the symbol @ denotes the channels of feature maps, while the number before the symbol denotes the size relative to the input image.

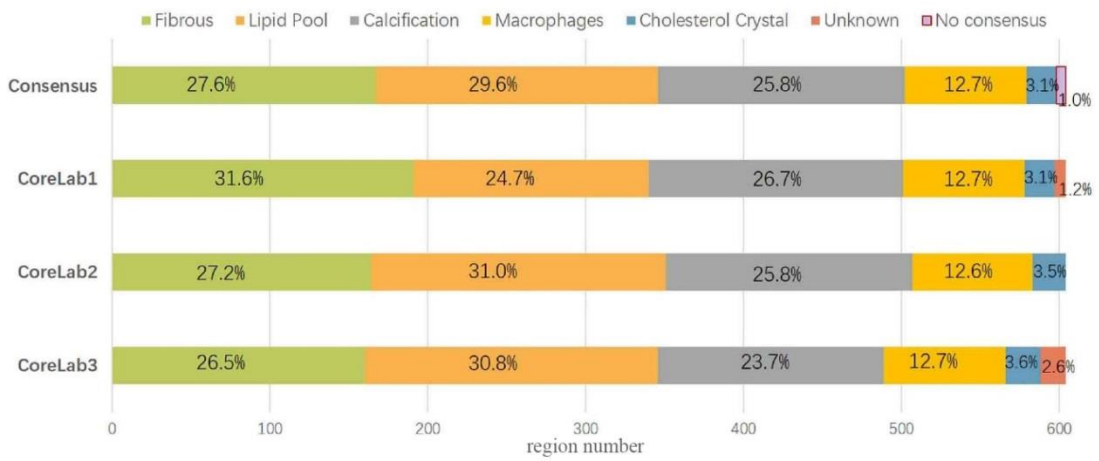


Supplementary Figure 2. Cases of plaque characterisation by core labs in the external validation data set.

A) Unanimity achieved in the three core labs.

B) Invisibility of media layer underlying thick fibrous cap evoked the controversy over the existence of lipid.

C) Atypically attenuated deposition confused the diagnosis.



Supplementary Figure 3. Consensus on plaque tissue types and agreement by the three core labs.

Supplementary Table 1. Ablation study results of all regions on the testing data set.

Feature bridges	Consecutive frames	Loss function	Average precision	Average recall	Average Dice
Horizontal	1	CEL	0.726	0.643	0.668
	3	CEL	0.731	0.637	0.663
	5	CEL	0.767	0.644	0.678
	7	CEL	0.765	**0.669**	**0.700**
	9	CEL	**0.778**	0.653	0.686
	11	CEL	0.747	0.643	0.671
	13	CEL	0.708	0.651	0.669
Horizontal	7	Combined	0.722	0.716	0.713
Horizontal+vertical	7	CEL	0.817	0.650	0.686
Horizontal+vertical	7	Combined	0.747	0.784	0.764

CEL: cross-entropy loss