Abstract

Objectives. Coronary angiography (CAG)-derived physiology methods have been developed in an attempt to simplify and increase the usage of coronary physiology, based mostly on dynamic fluid computational algorithms. We aimed to develop a different approach based on artificial intelligence methods, which has seldom been explored.

Methods. Consecutive patients undergoing invasive instantaneous free-wave ratio (iFR) measurements were included. We developed artificial intelligence (AI) models capable of classifying target lesions as positive (iFR ≤ 0.89) or negative (iFR > 0.89). The predictions were then compared to the true measurements.

Results. Two hundred-fifty measurements were included, and 3 models were developed. Model 3 had the best overall performance: accuracy, negative predictive value (NPV), positive predictive value (PPV), sensitivity, and specificity were 69%, 88%, 44%, 74%, and 67%, respectively. Performance differed per target vessel. For the left anterior descending artery (LAD), model 3 had the highest accuracy (66%), while model 2 the highest NPV (86%) and sensitivity (91%). PPV was always low/modest. Model 1 had the highest specificity (68%). For the right coronary artery, model 1 had an accuracy of 86%, NPV was 97%, and specificity was 87%, but all models had low PPV (maximum 25%) and low/modest sensitivity (maximum 60%). For the circumflex, model 1 performed best: accuracy, NPV, PPV, sensitivity, and specificity were 69%, 96%, 24%, 80%, and 68%, respectively.

Conclusions. We developed 3 AI models capable of binary iFR estimation from CAG images. Despite modest accuracy, the consistently high NPV is of potential clinical significance, as it would enable avoidance of further invasive maneuvers after CAG. This pivotal study offers proof of concept for further development.

Introduction

The use of invasive coronary physiology has been extensively studied and is clearly recommended in clinical guidelines today.[1,2] The most widely studied index is the Fractional Flow Reserve (FFR). Three major trials established its use in selecting lesions where revascularization had an additional benefit to medical therapy,[3-5] alongside large observational data.[6] More recently, another index gained

importance: instantaneous free-wave ratio (iFR). It was initially studied using FFR as the gold standard, with high accuracy.7,8 Our own experience showed similar results.9 Compared to FFR, iFR achieved non-inferiority in 2 major outcome trials,10,11 including over a 5-year follow-up period.12

Coronary physiology is vastly underused, ranging from 7% to 13% of procedures.13,14 The risk of complications, time consumption, and cost are likely reasons for this. Thus, calculating a physiological index (either FFR and/or iFR) digitally from coronary angiography (CAG) images is desirable, as it would bypass these limitations and potentially broaden physiology adoption. While this has already been achieved with several different software approaches relying on 3-dimensional (3D) vessel reconstruction and complex fluid dynamics computational algorithms,15-17 some limitations remain.

In medicine, artificial intelligence (AI) has shown great potential, especially in imaging, as several publications have demonstrated excellent results with regards to electrocardiogram (ECG),18 echocardiography,19,20 and magnetic resonance imaging (MRI).21,22 However, the use of AI regarding physiology estimation derived from CAG has seldom been explored.23,24 The potential advantages of using AI for this task could be either fully automating the process with minimal user input and/or improving the reliability of current systems, either functioning as a standalone approach or an added layer to existing software.

In this pilot study, we aimed to develop fully automated AI models capable of binary iFR lesion classification from CAG images alone, using measured invasive iFR as a reference.

Methods

Inclusion criteria. We conducted a single-center retrospective selection of consecutive patients over a 3-year period (2017-2019), who had undergone both CAG and invasive physiology assessment with iFR (Philips Volcano System), regardless of clinical context (ie, both acute and chronic coronary syndrome).

Exclusion criteria. We excluded cases where any of the following applied:

Imaging criteria:

Patients with cardiac devices or other sources of potential imaging artifacts overlapping with the coronary tree image

Poor image quality

Unsuccessful segmentation with AI models

Unclear individualization of lesion outline with overlapping vessels

Clinical criteria

History of coronary artery bypass grafting (CABG) or valvular intervention (surgical or percutaneous)

Significant left heart valvular disease (severe aortic stenosis or regurgitation, severe mitral regurgitation, moderate or severe mitral stenosis, moderate valvular disfunction in both aortic and mitral valves)

Target culprit vessel culprit of acute coronary syndrome (ACS)

Target vessel non-culprit of ACS in patient with ST-segment elevation myocardial infarction (STEMI) within 48 hours of presentation

Previous transmural myocardial infarction in target-vessel

Chronic total occlusion not previously treated with percutaneous coronary intervention (PCI) in any vessel

Left ventricular systolic disfunction, defined as an ejection fraction < 50%

Cardiogenic shock

Hemodynamic instability

Left main lesions

All the exclusion criteria were selected because of their potential impact on physiology index measurements, which could generate confoundment in the training process of an AI algorithm.

AI models development. We have previously trained AI models capable of fully automatic AI segmentation of CAG images.25-28 We used these models to segment all the CAGs of included patients. Images were then annotated with the target vessel and the location of the pressor sensor wires used to measure iFR using a single telediastolic frame where the target vessel was outlined. All available projections were annotated for each case.

Preliminary testing for AI models showed that performance was better when a single image that best outlined the target vessel was used, rather than using a combination of different projections. As a result, we proceeded to train our models as such. The model was given only the CAG image, both original (greyscale) and manually annotated after automatic segmentation (Figure 1). No further details, clinical or otherwise (such as the target vessel), were provided. The models were trained to binarily classify whether a given target had an iFR less than or equal to 0.89 or greater than 0.89, henceforth defined as a positive or negative iFR result, respectively.

Figure 1. Original diagnostic CAG image

Figure 1. Original diagnostic CAG image (left), automatic AI-based segmentation (center) and annotated image with coronary vessel proximal to wire sensor position in green and distal in blue (right). AI = artificial intelligence; CAG = coronary angiography.

A total of 3 models were trained. The first model used as input the sequence of diameters along the main branch of the analyzed vessel, automatically computed by a preprocessing algorithm. This sequence was then processed by a transformer encoder,[29] with a classification head on top, which inherently took into account the sequential nature of the data and allowed prediction of the iFR value at any given point within the artery. During training, the loss wa only computed at the point for which the ground-truth iFR was available, using a Cross-Entropy Loss, weighted by the inverse of each class's frequency. The Transformer encoder has 6 layers, 8 heads, a hidden dimension of 768, Gaussian Error Linear Unit (GELU) activation, 0.3 dropout, a maximum of 1024 linear positional embeddings, and is followed by a linear classification layer with 0.3 dropout.

Models 2 and 3 are Convolutional Neural Networks (CNNs), which took as input the concatenation of the single-channel angiography image and its segmentation, with 1 channel per class. Model 2 used a simple Cross-Entropy Loss, and model 3 used a Cross-Entropy Loss weighted by the inverse of each iFR class's frequency, aimed to mitigate the negative effects of class imbalance. The chosen CNN was an EfficientNet-B5,[30] with which we had already had success in previous work,[25–28] followed by a linear classification layer.

Theory suggests model 1 should be much more suitable for this task. Unlike model 1, the CNNs in models 2 and 3 cannot inherently consider the 1-dimensional characteristic of the artery and must learn to do so during training, possibly requiring more data to achieve the same level of performance. Additionally, the transformer in model 1 operates on a much lower-dimension input space than CNNs, making the former's task theoretically much easier. Finally, since the location of the iFR predicted by models 2 and 3 is directly tied to the input segmentation, they require an additional inference per iFR prediction.

Performance assessment and statistical analysis. Descriptive variables are shown in absolute and relative (percentage) numbers. Quantitative variables are shown in average ± standard deviation (if normally distributed) or median (interquartile range) if non-normally distributed. The chi-square test was used for statistically comparing the binary classification of measured iFR vs that of the models. A P-value of .05 was used for statistical significance.

The results of the models' classification of target lesions as either iFR positive or negative were compared to those of the true (ie, real) invasive iFR measurements, as follows:

True positive (TP): both estimated and real iFR were positive

False positive (FP): positive estimated iFR and negative real iFR

True negative (TN): both estimated and real iFR were negative

False negative (FN): negative estimated iFR and positive real iFR

Using this classification, the following parameters were calculated:

Accuracy: ([TP + TN] / [TP + TN + FP + FN])

Sensitivity: TP / (TP + FN)

Specificity: TN / (TN + FP),

Positive predictive value (PPV): TP / (TP + FP)

Negative predictive value (NPV): TN / (TN + FN)

To ensure proper evaluation of the results, the models could not be tested on data already seen during training. Hence, we used a cross-validation split at the patient level into 10 subsets, retaining the relative distribution of target vessel and iFR classification per split. The models' iFR classification in each subset was then undertaken using neural networks trained exclusively on the remaining data. This enabled the assessment of the models' performance for the whole cohort, whereas the usual splitting in a fixed train/test datasets would have resulted in a much smaller testing group, limiting the ability to test the models' performance. We have successfully used this approach in the past, when developing our segmentation models.25-27 SPSS 27 (IBM) was used for analysis.

Ethical issues. This study complies with the Declaration of Helsinki and was approved by the local Ethics' Institutional Review Board.

Results

Baseline characteristics. A total of 334 patients were screened. After applying the exclusion criteria, a total of 250 measurements, from a total of 223 patients, were included (Figure 2, Table 1). Most lesions had an iFR greater than 0.89. There was a large imbalance between positive and negative iFR lesions in the right and circumflex coronary arteries subgroups. The difference was much less pronounced in the LAD measurements (Table 2).

Figure 2. Inclusion flowchart.

Figure 2. Inclusion flowchart. iFR = instantaneous free-wave ratio.

Table 1. Clinical characteristics

Table 2. iFR results

Physiology Prediction of AI Models

Overall results. The difference between measured iFR classification and that of the AI models was not statistically significant for model 1 (P = .063), whereas models 2 and 3 differed significantly (P < .001). Models 1 and 3 classified most lesions as negative, contrary to model 2, with only model 1 distributing classifications in similar proportions to those of the ground truth. Agreement was strongest for negative lesions. Details are presented in Table 3.

Table 3. iFR classifications

All models performed with modest accuracy, with the worst results for model 2 (58%) and the best for model 3 (nearing 70%), followed closely by model 1 (65%). NPV was high for all models, performing close to 80% or 90%, in contrast with PPV. Models 2 and 3 had the highest sensitivity, whereas model 1 had the highest specificity. Details are presented in Figure 3.

Figure 3. Performance metrics

Figure 3. Performance metrics (with 95% CI) of each AI model regarding iFR binary classification using the measured iFR classification as reference, for all iFR measurements across all target vessels. Accuracy, PPV, NPV, sensitivity, and specificity are displayed. Background in green for values close to 100%, red for values approaching 0% and mixed (yellow) for mid-range values. AI = artificial intelligence; iFR = instantaneous free-wave ratio; NPV = negative predictive value; PPV = positive predictive value.

Left anterior descending (LAD) lesions. The models' classification of iFR as compared to the measured iFR was not statistically significant for model 1 (P = .854), whereas models 2 and 3 differed significantly (P < .001). Models 1 and 2 classified most lesions as negative, where agreement was more common. Model 3 classified most lesions as positive. Details are presented in Table 4.

Table 4. iFR Classifications

Models 2 and 3 displayed modest accuracy, with better results for the latter, while model 1's performance barely surpassed 50%. The NPV and sensitivity was high or very high for model 2, nearing 90%, while model 3's performance was 78% for both. Model 1 performed reasonably only for specificity (68%). The PPV was above 50% for models 2 and 3. Details are presented in Figure 4.

Figure 4. Performance metrics

Figure 4. Performance metrics (with 95% CI) of each AI model regarding iFR binary classification prediction using the measured iFR classification as reference, for iFR measurements in the LAD. Accuracy, PPV, NPV, sensitivity, and specificity are displayed. Background in green for values close to 100%, red for values approaching 0% and mixed (yellow) for mid-range values. AI = artificial intelligence; iFR = instantaneous free-wave ratio; LAD = left anterior descending artery; NPV = negative predictive value; PPV = positive predictive value.

Right coronary artery (RCA) lesions. The iFR classification of model 1 differed significantly from the measured iFR (P = .005), whereas for models 2 and 3 there were no significant differences (P = .282 and .357, respectively). All models classified most lesions as negative (albeit in smaller proportion to the actual measurements distribution), especially model 1. Agreement was highest for negative lesions in all models. Details are presented in Table 5.

Table 5. IFR Classifications

Models 1 and 3 displayed the highest accuracy, especially the former (86%), which was in contrast with model 2. The NPV was always very high, with a maximum of 97% for model 1. On the opposite spectrum of performance, the PPV was very low for all models. Sensitivity was high for both models 1 and 3. Lastly, specificity was low for models 2 and 3, and modest for model 1. Details are presented in Figure 5.

Figure 5. Performance metrics

Figure 5. Performance metrics (with 95% CI) of each AI model regarding iFR binary classification prediction using the measured iFR classification as reference, for iFR measurements in the RCA. Accuracy, PPV, NPV, sensitivity, and specificity are displayed. Background in green for values close to 100%, red for values approaching 0% and mixed (yellow) for mid-range values. AI = artificial intelligence; iFR = instantaneous free-wave ratio; LAD = left anterior descending artery; NPV = negative predictive value; PPV = positive predictive value; RCA = right coronary artery.

Circumflex artery (Cx) lesions. The neasured iFR classification differed significantly from the iFR classification of model 1 (P = .039), but not models 2 and 3 (P < .673 and .333, respectively). All models classified most lesions as negative, but always in quite shorter proportion to the actual measurements' distribution. Agreement was highest for negative lesions in all models. Details are presented in Table 6.

Table 6. IFR Classifications

Model 1 scored higher in all metrics, with an accuracy of 69% for a very high NPV (96%) and high specificity (80%). Sensitivity was 68% and the PPV was low (24%). Model 3 followed with similar, albeit inferior performance. Model 2 only scored high in NPV (87%). Details are presented in Figure 6.

Figure 6. Performance metrics

Figure 6. Performance metrics (with 95% CI) of each AI model regarding iFR binary classification prediction using the measured iFR classification as reference, for iFR measurements in the Cx. Accuracy, PPV, NPV, sensitivity, and specificity are displayed. Background in green for values close to 100%, red for values approaching 0% and mixed (yellow) for mid-range values. AI = artificial intelligence; Cx = circumflex artery; iFR = instantaneous free-wave ratio; LAD = left anterior descending artery; NPV = negative predictive value; PPV = positive predictive value; RCA = right coronary artery.

Discussion

Main findings: a proof of concept. In this study, we were able to develop AI models capable of binary iFR lesion classification. The accuracy of all models was modest, close to 70% for model 3. For both the LAD and Cx, the best that any model could achieve was also just below 70%. However, for the RCA, an accuracy of 78% and 86% was achieved for models 3 and 1, respectively, which, in the latter case, is quite high.

The models displayed very different reliability in correctly classifying lesions as either positive (ie, iFR ≤ 0.89) or negative (ie, iFR > 0.89). Indeed, the PPV was low or very low for all models, likely because much fewer positive cases were available for training. Only for the LAD, where there was a larger number of positive lesions, were the models able to achieve modest performance, surpassing the 50% PPV mark. The NPV, however, was generally high or very high, ranging from 77% to 89% overall. For both the RCA and the Cx, at least 1 model neared 100%.

At first glance, one might interpret these findings to be a result of the predominance of negative iFR cases, overwhelmingly so for the RCA and the Cx. Therefore, the models could simply be producing a result based on the statistically higher likelihood of a negative result, rather than extrapolating from CAG images.

We believe several factors suggests otherwise. First, the models did not necessarily reproduce the distribution of positive/negative lesions of the true measurements, neither globally nor regarding specific vessels. For example, model 3, which achieved the highest accuracy overall, only classified lesions as negative in 56% of cases (vs 74% for real measurements). The difference was even greater for the Cx, where the models' negatively classified lesions ranged from 51% to 62% (vs a real result of 89%).

Furthermore, the models' NPV was also greater than the proportion of negative cases, reaching 86% for the LAD (vs 54% real negative cases) and very close to 100% for both the RCA and Cx (where the real proportion of negative cases was close to 90%). Thus, the number of iFR-negative cases provided enough training data for the models to correctly learn to classify a lesion as negative with high reliability.

Despite this, a significant number of true negative lesions were not identified, as evidenced by modest specificity overall. Models 1 and 3 were exceptions in the case of RCA lesions, where higher performances were obtained.

No single model proved to be ideal. Model 1, followed by model 3, seemed to be the best option for the RCA and the Cx, whereas both models 2 and 3 were a better fit for the LAD. This suggests that fine-tuning a model for the target vessel may improve results.

Considering all of the above, we believe our models offer proof of concept that the derivation of coronary physiology data by deep learning AI models based on X-ray angiography alone is feasible.

Potential practical clinical implications. When faced with either a positive or negative result by any such software, the operator's main question is often whether the result is likely correct (ie, how high is the PPV and/or NPV). Arguably, in the context of invasive coronary physiology, the NPV is of particular importance: a negative result enables the operator to conclude the procedure without engaging in further invasive maneuvers (ie, deploying a guide catheter, wiring the target vessel, administering further drugs), as opposed to a positive result. The fact that the majority of measurements in invasive physiology are negative further strengthens this point.[9–11,15,17,23,31] For example, in this case series, 185 (74%) cases had an iFR greater than 0.89. Since the models classified lesions as negative in 42% to 72% of cases and the NPV approaches 89%, in the best-case scenario, arguably around one- to two-thirds of all measurements could have been avoided. For the RCA and Cx, where the NPVs neared 100% and the proportion of cases classified as negative was even higher, the impact would have been even more significant.

Thus, while the models are by no means reliable enough for current clinical deployment, this somewhat simplified analysis illustrates the potential practical implications of this technology.

Analyzing lesions on coronary angiography using AI: other studies. Two studies of FFR estimation from CAG using primarily AI methods were published. However, to the best of our knowledge, ours is the first published study to report fully automatic derivation of iFR from CAG using such methods.

Roguin et al23 conducted a pilot feasibility study in a single-center population consisting of 31 patients with predominantly LAD lesions (80%). They reported an accuracy of 90%, an NPV of 87%, a PPV of 94%, sensitivity of 88%, and specificity of 93% when conducting a binary analysis of FFR ≤ 0.80, similar to our approach. Their single model is able to derive an estimated FFR value, with an area under the curve of 0.91 and an r correlation coefficient of 0.71 (P < .001). The study does not report the exact AI methods. However, the binary approach to classify lesions, fully based on AI, using routine angiography projections (rather than predefined angulations) with a fully automatic method, is conceptually very similar to ours. While these results are impressive, with a seemingly superior performance compared to ours, the small sample size and single center nature of the study are limiting factors.

Cho et al24 used a different approach. With a very large sample of 1501 lesions from a single center (predominantly LAD [67%]), the authors plotted the target vessel diameters together with clinical characteristics (age, sex, body surface area, and target segment) to binarily classify FFR measurements with a threshold of less than or equal to 0.80. An overall accuracy of 82%, an NPV of 84%, a PPV of 81%, sensitivity of 84%, and specificity of 80% were reported in the test set, similar to the external validation dataset of 79 patients. The main limitation of this study, from an AI research perspective, is that it does not provide a fully automated AI physiology estimation, as it requires manual segmentation and diameter calculations of the vessel with an external software, thereby rendering the process semi-automatic and somewhat time-consuming.

Another group is currently launching an initiative for fully automated AI-based PCI guidance interpretation, including FFR32, but the project is underway and results are not yet available.

Lastly, one of the largest groups in iFR research has applied AI for interpreting iFR pullback curves and found non-inferiority to human performance.33 While this is a very different task from the one we explored, it highlights how AI is also being applied to this important index which is so commonly used in clinical practice.

There has also been some exploration of the estimation of Coronary Flow Reserve (CFR) from CAG using AI methods, mainly to perform microcirculatory studies.34 Other authors have tested the application of deep learning methods to classifying stenosis using Quantitative Coronary Analysis (QCA)35 or automatically detecting significant coronary stenosis using bounding boxes,36,37 which could be of use given the operators' heterogeneity and tendency for lesion overestimation when interpreting CAG, as we have recently demonstrated ourselves.28

Physiology derivation from CAG images without a primary AI approach: other studies. The estimation of physiology from CAG images has been explored in recent years, with commercial software made recently available.

Most studies focused on FFR, using a threshold of less than or equal to 0.80. The FAST-FFR pivotal trial was a multicenter international study of 301 subjects with a predominance of LAD lesions (54.2%). The correlation between estimated (FFRangio) and measured FFR was r = 0.80 (P < .001), with an accuracy of 92.2%, an NPV of 94.8%, a PPV of 89%, sensitivity of 93.5%, and specificity of 91.2%.17 A pooled analysis of 5 cohort studies yielded similar results.38

The FAST I31 and II16 studies tested a similar approach. In the larger FAST II study16, 334 patients from 6 centers were enrolled (with 66% LAD lesions). The correlation with invasive FFR was r = 0.74 (P < .001). Two multicenter trials are ongoing to test the clinical outcomes of virtual (vFFR) vs invasive FFR approach.39,40

The Quantitative Flow Ratio (QFR) is perhaps the most extensively "virtual" FFR index studied to date. After the encouraging results of the multicenter FAVOR pilot study (73 patients, 84 vessels, mostly LAD [54.8%]),41 2 further studies enrolled over 600 patients from Europe and Asia.15,42 In the one with a large European cohort, values of 86.8%, 93%, 76.3%, 86.5%, and 86.9% for accuracy, NPV, PPV, sensitivity, and specificity were obtained, respectively, with a correlation of r = 0.83 (P < .001).15 A Chinese multicentric trial comparing a PCI QFR-guided strategy vs an invasive FFR-guided strategy yielded better outcomes for the QFR group.43

The derivation of iFR from CAG has also been explored on the REVEAL iFR trial.44 Published results are expected soon.

All of the above studies employ primarily non-AI methods to derive FFR from CAG, using a combination of 3D-image reconstruction and computational fluid dynamics. They demonstrate that physiology can successfully be derived from CAG images alone and have a meaningful impact on clinical outcomes. Further ongoing research is likely to strengthen this approach. While all of the above post-pilot studies

yielded better performance results than our model, one thing seems ubiquitous: the NPV is always high, which is of particular significance.

However, these approaches are not without disadvantages. A reasonable amount of manual input is necessary – marking the target vessel, proximal and distal points, defining "healthy" regions, or correcting for imperfections in manual segmentation – as we have experienced ourselves when testing these platforms at our own catheterization laboratory. They are therefore semi-automatic, somewhat time consuming, and may potentially not be as reliable in less experienced hands. The FAST II trial clearly illustrated this, as the performance at specific sites was lower than at the core lab; the reported overall accuracy, NPV, PPV, specificity, and sensitivity were 83% vs 90%, 85% vs 90%, 79% vs 90%, 71% vs 81%, and 89% vs 95%, respectively.16

Lastly, except for one,44 the above-mentioned methods require more than 1 projection (sometimes prespecified), which is probably the result of multiple factors. Indeed, the 3D nature of the coronary anatomy and the existence of energy losses in very distal segments (with resulting lower pressures), along with the limitations of CAG resolution and motion artifacts, all render a 3D approach more reliable, and are likely playing a role in hampering our models' accuracy, since it was based on a single 2-dimensional (2D) frame. Thus, while our simplified approach may initially be perceived as advantageous, it was, in all likelihood, a limitation.

Limitations and future directions. As AI training is highly volume dependent, the dataset size was the most important limitation. This was especially relevant for cases with iFR less than or equal to 0.89 and cases pertaining to the Cx. However, our relative distribution regarding target vessel and positive/negative cases is in agreement with previously published datasets.10,11,15,17,23,31 As a result, obtaining a dataset with enough iFR-positive cases for successful training, especially concerning the RCA and Cx, will require a much larger dataset. The dataset size was also limiting with regards to the method employed for model testing. Using a classical train/test split of 80% / 20% would have resulted in a small testing dataset and reduced our ability to test the models' performance, particularly for subanalyzing performance per target vessel, given the naturally unbalanced characteristics of the dataset. We therefore preferred a cross-validation split in 10 subsets, as mentioned in the Methods section. While this is a common approach in the field of machine learning, it may be regarded as a limitation as well.

Some may view the extrapolation of iFR rather than FFR as a limitation, because FFR was directly compared to angiography in clinical outcomes trials,3-5 whereas iFR was only directly compared to FFR itself10-12 However, iFR has repeatedly been shown to be non-inferior to FFR and today is (together with other resting indexes) the default tool of epicardial physiology assessment in many labs due to its simplicity. As a result, the current amount of iFR measurements far outpaces those of FFR in our lab, and thus iFR provides a much larger base for future training, improvement, and validation.

Another limitation is the fact that the model provides a binary classification, but not yet the iFR value itself. During preliminary testing, it was clear that determining the exact iFR value would require a much larger dataset, which was beyond the scope of a pivotal study.

The use of a single image (human-optimized and labelled end-diastolic 2D frame) for the models' training, instead of a 3D reconstruction based on multiple 2D projections, was also a limitation, given the truly 3D nature of coronary anatomy and lesions. The fact that the above-mentioned non-AI approaches obtained a superior performance with 3D reconstruction, rather than a 2D approach like ours, further supports this consideration.

The single-center, retrospective dataset is another limitation, as external validation will be required in the future. Notwithstanding, physiology results have been shown to be quite reproducible and thus the impact of this particular limitation is not very likely to be of significance.

The concept of this study is thus exploratory and aimed at proof of concept. We aim to greatly enlarge the training dataset through multi-institutional collaboration (as we have done for our segmentation models[27,28]), which will be essential for performance improvements and external validation. We also aim to further improve our models using multiple projections and 3D reconstruction, which may enhance performance. The ultimate aim is deployment in clinical practice; this may occur either as a standalone solution, or by improving existing software.

Conclusions

We developed deep learning AI models capable of binary lesion classification using an iFR threshold of 0.89. While the overall accuracy of the models is not yet high enough for clinical deployment, the high negative predictive capacity of these models is of clinical significance and potential clinical application. This pivotal study therefore offers proof of concept for further development, with larger and multicentric training and validation datasets poised for the future. This approach has the potential to evolve into a standalone software aid in the catheterization laboratory, or for further integration into existing, non-AI based software solutions by streamlining workflows and/or improving their performance. This could prove to be of great value for patient management and to improve catheterization laboratory flows.

Affiliations and Disclosures

From the 1Structural and Coronary Heart Disease Unit, Cardiovascular Center of the University of Lisbon (CCUL@RISE), Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal; 2Serviço de Cardiologia, Departamento de Coração e Vasos, CHULN Hospital de Santa Maria, Lisbon, Portugal; 3INESC-ID, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal; 4Neuralshift Inc., Lisboa, Portugal; 5Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal.

Address for correspondence: Miguel Nobre Menezes, MD, MSc, Serviço de Cardiologia, Avenida Professor Egas Moniz, Lisbon, 1649-028, Portugal. Email: mnmenezes.gm@gmail.com; X: @nobremenezes