IJC Heart & Vasculature 55 (2024) 101528

Contents lists available at ScienceDirect

IJC Heart & Vasculature

journal homepage: www.sciencedirect.com/journal/ijc-heart-and-vasculature

Evaluating machine learning accuracy in detecting significant coronary stenosis using CCTA-derived fractional flow reserve: Meta-analysis and systematic review

Danny van Noort [a,1,*], Liang Guo [b,c,1], Shuang Leng [a], Luming Shi [b,c,d], Ru-San Tan [a,d], Lynette Teo [e,f], Min Sen Yew [g], Lohendran Baskaran [a,d], Ping Chai [f,h], Felix Keng [a,d], Mark Chan [f,h], Terrance Chua [a,d], Swee Yaw Tan [a,d], Liang Zhong [a,d]

[a] NHRIS, National Heart Centre, Singapore
[b] Singapore Clinical Research Institute, Consortium for Clinical Research and Innovation, Singapore
[c] Cochrane, Singapore
[d] Duke-NUS Medical School, Singapore
[e] Department of Diagnostic Imaging, National University Hospital, Singapore
[f] Yong Loo Lin School of Medicine, National University Hospital, Singapore
[g] Tan Tock Seng Hospital, Singapore
[h] Department of Cardiology, National University Hospital, Singapore

ARTICLE INFO

Keywords:
Machine learning
Coronary computed tomography angiography (CCTA)
Fractional flow reserve (FFR)
Invasive coronary angiography (ICA)

ABSTRACT

Background: The use of machine learning (ML) based coronary computed tomography angiography (CCTA) derived fractional flow reserve (ML-FFR$_{CT}$), shortens the time of diagnosis of ischemia considerably and eliminates unnecessary invasive procedures, when compared to invasive coronary angiography with invasive FFR (iFFR). This systematic review aims to summarize the current evidence on the diagnostic accuracy of (ML-FFR$_{CT}$) compared with iFFR for diagnosis of patient- and vessel-level coronary ischemia.

Methods: To identify suitable studies, comprehensive literature search was performed in PubMed, the Cochrane Library, Embase, up to August 2023. The index test was ML derived FFR and studies with diagnostic test accuracy data of ML-FFR$_{CT}$ at a threshold of 0.8 were included for the review and *meta*-analysis. Quality of evidence was assessed using QUADAS-2 checklist.

Results: After full text review of 230 identified studies, 17 were included for analysis, which encompassed 3255 participants (age 62.0 ± 3.7). 8 studies reported patient-level data; and 12, vessel-level data. With iFFR as the reference standard, the pooled patient-level sensitivity, specificity, and area-under-curve (AUC) of ML-FFR$_{CT}$ were 0.86 [95 % CI: 0.79, 0.91], 0.87 [95 % CI: 0.76, 0.94], and 0.92 [95 % CI: 0.89–0.94], respectively; and pooled vessel-level sensitivity, specificity, and AUC, 0.80 [95 % CI: 0.74–0.84], 0.84 [95 % CI: 0.77–0.89), and 0.88 [95 % CI: 0.85–0.91], respectively.

Conclusions: This systemic review demonstrated the favourable diagnostic performance of ML-FFR$_{CT}$ against standard iFFR, although heterogeneity exists, providing support for the use of ML-FFR$_{CT}$ as a triage tool for non-invasive screening of coronary ischemia in the clinical setting.

Abbreviations: Acc, Accuracy; AI, artificial intelligence; AUC, area under the curve; BRNN, bi-directional recursive neural network; CAD, coronary artery disease; CCTA, coronary computed tomography angiography; CFD, computational fluid dynamics; CFD-FFR$_{CT}$, computational fluid dynamics based fractional flow reserve derived from computed tomography angiography; CTCA, computed tomography coronary angiography; DS, diameter stenosis; FFR, fractional flow reserve; ICA, invasive coronary angiography; iFFR, invasive fractional flow reserve; ML, machine learning; ML-FFR$_{CT}$, machine learning-based fractional flow reserve derived from computed tomography angiography; MLNN, multilevel neural network; FN, false negative; FP, false positive; NPV, negative predictive value; PPV, positive predictive value; QUADAS-2, quality assessment of diagnostic accuracy studies, version 2; RoB, risk of bias; Sn, sensitivity; Sp, specificity; TN, true negative; TP, true positive.

* Corresponding author at: National Heart Centre Singapore, 5 Hospital Drive, 169609 Singapore.
E-mail address: danny.van.noort@nhcs.com.sg (D. van Noort).

[1] shared first authors.

https://doi.org/10.1016/j.ijcha.2024.101528
Received 12 May 2024; Received in revised form 30 September 2024; Accepted 7 October 2024
Available online 18 October 2024

1. Introduction

Coronary artery disease (CAD) is the most common cause of cardiovascular mortality [1]. It is caused by atherosclerotic build-up of plaque in the walls of coronary arteries, which results in stenosis of the coronary lumen and compromised blood flow to the heart muscle. The reference standard for determining the anatomical severity of the coronary lesion is binary diameter stenosis (DS) determined at invasive coronary angiograph (ICA); and its physiological significance, invasive fractional flow reserve (iFFR). iFFR is the ratio of pressures invasively measured proximal and distal to the coronary lesion during ICA, under conditions of parenterally administered vasodilator-induced maximal hyperemia [2]. FFR $\leq$ 0.8 indicates ischemia caused by the corresponding coronary lesion [3].

iFFR measurement during ICA is guideline-recommended for the evaluation of coronary stenosis before coronary revascularization [4,5], as it has been shown to safely reduce unnecessary revascularizations compared with DS-driven approaches. The disadvantages of iFFR are the risks and costs associated with additional pressure wire instrumentation and hemodynamically active vasodilator drug administration. iFFR-related complications include local vascular injury, myocardial infarction, cerebrovascular complications, and even death [6].

Coronary computed tomography angiography (CCTA) allows non-invasive assessment of CAD plaque burden and plaque characteristics. It is widely used to diagnose CAD and select patients for ICA, with a view to coronary revascularization. The determination of the extent of calcified/non-calcified coronary plaques and identification of high-risk plaque characteristics rely largely on visual assessment by experts [7], which often overestimate the degree of anatomical stenosis [8,9]. Analogous to iFFR, it is possible to assess the physiological significance of coronary lesions on CCTA. Using computational fluid dynamics (CFD) to calculate local blood pressures at any and all luminal locations along the three-dimensional reconstructed coronary tree with simulated hyperemia assumptions, FFR can be derived from patient-specific CCTA image data (FFR_{CT}) [10] and inputs, including systolic and diastolic blood pressures, and CCTA-calculated left ventricular mass [11]. As CFD relies on formal grounds-up mathematical simulation, the processing and calculation time is onerous. Even with a more efficient reduced-order CFD simulation, it can take up to 2 h, depending on the complexity of the patient-specific coronary anatomy and computer processing speed [12]. An alternative faster approach would be to train machine learning (ML) models directly on datasets containing patient-specific clinical information and CCTA images with prior CFD-derived FFR_{CT} annotations. This ML-based FFR_{CT} obviates the need for post-acquisition *de novo* CFD, significantly shortening its turnaround time, which may help garner clinical adoption. This systematic review aims to summarize the current evidence on the diagnostic accuracy of (ML-FFR_{CT}) compared with iFFR for diagnosis of patient- and vessel-level coronary ischemia.

2. Methods

2.1. Information sources and search strategy

This systematic review has been registered in the international prospective register of systematic reviews (PROSPERO) under the registration number: CRD42023458316. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were adopted to identify eligible studies and conduct the whole review. A systematic search was performed in PubMed, Embase and the Cochrane Library for suitable publications published in any language up to August 2023. The following search terms were used: artificial intelligence, machine learning, factional flow reserve, FFR, angiography, stenosis and related synonyms. The search included both MeSH terms and free text. In addition, all reference lists in the included articles and relevant review articles were manually checked. The studies were all independently reviewed by 2 reviewers (DvN, LS) and any discrepancy was resolved by a third reviewer (ZL). All retrospective and prospective studies that reported diagnostic accuracy of machine learning FFR and also a standard invasive coronary angiography with invasive FFR reported, were considered eligible. Review articles, case reports, comments, poster presentations, and studies without complete data on diagnostic accuracy were excluded. From the final enrolled abstracts, full-text articles were retrieved and evaluated.

2.2. Selection process and data collection

Studies were selected based on population, index test, reference test, and target condition. The inclusion criteria were: (1) patients underwent CCTA and ICA scans for stenosis evaluation; (2) FFR was measured by ICA and iFFR $\leq$ 0.8 was diagnosed as ischemia or hemodynamically significant, which was taken as the reference standard for this review; (3) ML-FFR_{CT}, the index test, was calculated by a machine learning algorithm and ML-FFR_{CT} $\leq$ 0.8 was considered as indicator of ischemia; (4) data was reported at patient and/or vessel level; and (5) studies reported endpoint data of interest, namely sensitivity (Sn), specificity (Sp), accuracy (Acc), positive predictive value (PPV) and negative predictive value (NPV) or true positives (TN), true negatives (TN), false positives (FP) and false negatives (FN).

Data was extracted by DvN and LG, independently, into a pre-defined data extraction form. Baseline characteristics consists of authors, year of publication, study design, patient demographics, vessel information, equipment information, type of ML algorithm used and outcomes. The accurate number of true-positive, true-negative, false-positive, and false-negative results was extracted to construct a 2x2 table from qualified studies for calculation on diagnostic accuracy parameters. Authors were contacted if relevant information was missing or wrongly reported.

2.3. Study risk of bias assessment

Risk of bias assessment (RoB) was independently conducted by DvN and LG using the Quality Assessment of Diagnostic Accuracy Studies, Version 2 (QUADAS-2), which comprised four key domains: patient selection, index test, reference standard, flow and timing. Each domain was assessed for RoB, and the first three were also evaluated for applicability. Signaling questions were included to inform judgments regarding the risk of bias: a domain would be rated as “high risk” if the response to at least one nested signaling question was “No”, rated as “unclear risk” if 1 or more unclear out of 2 signaling questions or 2 or more unclear out of 3 questions. Any disagreement in quality assessment was resolved via consensus. A study was classified as having a high RoB if at least one of the domains of QUADAS-2 was rates as “high risk”.

2.4. Statistical methods

Summary receiver operating characteristic (SROC) curves were generated to illustrate the overall diagnostic accuracy of the index test (Machine Learning FFR_{CT}) compared with the reference test (invasive FFR). A bivariate random-effects model was used for *meta*-analysis to calculate summary estimations of Sn, Sp, likelihood ratios (LR), the diagnostic odds ratio (DOR), and areas under the SROC curve (AUC). Forest plots of Sn and Sp were presented for studies with the common threshold. Fagan’s nomograms were used to assess the relationship between the pre-and post-test probability of the disease and the LR of the diagnostic test. Sensitivity analysis was implemented to identify potential sources of heterogeneity. Non-threshold heterogeneity was evaluated by the Q test and I^2, with a threshold of p-value < 0.1 and $I^2 >$ 50 % suggesting an obvious heterogeneity respectively. RoB assessment was conducted in RevMan 5.3. Statistical analyses were conducted in RevMan 5.3 and STATA 17 with Midas and Metandi package.

3. Results

3.1. Study selection and characteristics

A total of 230 articles were identified in our literature search, and 98 eligible full-text articles were further reviewed after duplicates removal and title/abstracts screening. Following the inclusion criteria, 17 studies were recruited into the systematic review and *meta*-analysis (Supplementary Figure S1). The studies came from 8 different countries (China, Germany, Japan, the Netherlands, Poland, South Korea, Sweden and USA), of which 2 prospective cohorts, 14 retrospective cohorts, and 1 mixed cohort (prospective and retrospective) (Table 1). Eight studies reported patient level data and 12 reported vessel level data.

3.2. Demographics

In total, 3255 patients with 3906 vessels from studies published from 2017 to 2022 were included in the analysis. The patients were recruited to these various studies from 2008 to 2019 and were mostly retrospective studies, with a mean age of 62.0 (± 3.7). The number of the patients of the various studies included in the analysis ranged from 63 to 484, while for vessels this was 71 to 618. The demographics for the different studies are presented in Table 1. From the 17 studies, 3 did not use any of the Siemen's Somatom CT-scanners [96, 97, 98], but an Aquilion scanner from Canon or Revolution CT, GE. These 3 studies, together with another study [41], didn't use the Siemens ML-algorithm either. Among a set of different ML algorithms that were used in included studies, only the best performers were retrieved for further analysis.

3.3. Quality assessment in included studies

The quality of all included studies was evaluated based on the QUADAS-2 protocol, assessing the RoB and applicability concerns (Supplementary Figure S2). In general, included studies performed well in the applicability concerns domain and had a low risk of bias on index test as all included studies reported ML-FFR_{CT}. Regarding the patient selection domain, more than half studies had an unclear risk due to lack of information on whether patients enrolled consecutively or randomly; 3 studies had high risk due to inappropriate exclusions. Regarding flow and timing, 1 study had high risk of bias due to multiply reference test methods used in the participants. Patients were included in the analysis or lack of information. Other unclear risks in the quality assessment were mainly due to inadequate information.

3.4. AI methods

The use of AI has become more prevalent in health research. However, in the field of FFR prediction, there seems to be, at the moment, only one major player, and that is an algorithm designed by Siemens (cFFR) which was reported in about 76 % of the publications. In this review the included studies covered 3 different ML methods.

The ML-based algorithm by Siemens Healthineers used 12,000 synthetic generated coronary geometries [13]. For each geometry the FFR_{CT} was calculated for the entire coronary tree using a reduced-order computational blood flow model. The ML-based algorithm was trained using a deep neural network with 4 hidden layers with 10,000 of these geometries. The algorithm was validated with the other 2,000 geometries.

DEEPVESSEL-FFR from the company Keya Medical was an alternative method used in 3 of studies included in this *meta*-analysis is. This algorithm uses deep learning architecture constructed with a multilevel neural network (MLNN) and a bi-directional recursive neural network (BRNN). During training, ground truths, generated by solving the Navier-Stokes equations with iFFR as a reference, were applied to the output layers.

Only 1 study [14] was fully automated, from segmentation to FFR. In all the other studies, segmentation was semi-automatic. This is also the only study using an ML-algorithm that was not from a company. It consisted of fully connected neural network comprising 256 nodes in one hidden layer and virtual adversarial training to predict the FFR value.

3.5. Data synthesis

Fig. 1 shows the forest plot of sensitivity and specificity from all studies. Data has been analysed at 2 levels; patient-level [14–20] and vessel-level [15–17,19–30]. The sensitivity and specificity of the ML-FFR_{CT} at patient-level ranged from 0.79 to 0.97 and 0.63–1, respectively, and the accuracy ranged from 0.76 to 0.99. Following bivariate random-effects model, the pooled sensitivity and specificity was 0.86 [0.79, 0.91] and 0.87 [0.76, 0.94], respectively, and the accuracy was 0.85 [0.78, 0.92] (Supplementary Table ST1). Pooled AUC was 0.92 [0.89–0.94].

The sensitivity and specificity of the ML-FFR_{CT} at vessel-level ranged from 0.61 to 0.98 and 0.63–0.97, respectively, with the accuracy ranged from 0.63 to 0.96. The pooled sensitivity and specificity were 0.8 [0.74, 0.84] and 0.84 [0.77, 0.89], respectively, and the accuracy was 0.81 [0.76, 0.86]. Pooled AUC at vessel level was 0.88 [0.85–0.91].

Summary receiver operating characteristic illustrated that patient-level data tended to have better performance when compared with vessel-level data (Fig. 2), which also echoed the summary estimates that patient-level studies reported higher AUROC. However, it seems vessel-level data from included studies was slightly less heterogenous as both confidence and prediction region were smaller compared to patient-level data.

Still, both level data indicated relatively high between-study heterogeneity based on I^2 (>90 %). To explore the potential source of heterogeneity, sensitivity analysis was conducted on key covariates. Generally, results were quite similar under different covariates which indicated robust results on our summary estimates though the I^2s were constantly high. It was only among patient-level studies with low or unclear risk of bias, the I^2 reduced dramatically to 58 %, though still high and with wide confidence interval, the quality of study could potentially contribute to the heterogeneity of the results. Meta-regression also showed similar result that RoB level could be potential source of heterogeneity on both sensitivity and specificity($P = 0.03$). Due to limited number of studies, subgroup analysis was not available in our case.

The positive likelihood ratio (LR +) and negative likelihood ratio (LR-) for ML-FFR_{CT} were 6.6 [3.4, 13.1] and 0.16 [0.10, 0.25] at patient-level, 5 [3.4, 7.4] and 0.24 [0.18, 0.32] at vessel-level, respectively, which indicated a moderate discrimination of the disease. Fagan's plots (Fig. 3a) illustrated that, at patient-level, with given pre-test probabilities of all types of heart disease prevalence among adults at about 11 % [31],or an estimated prevalence of ischemic disease among patients who visit CVD clinics (based on clinical experience) at 50 %, a significant ML-FFR_{CT} result (<=0.8) revealed incremental probabilities of iFFR stenosis to 45 % and 87 %. Conversely, an absence of significant ML-FFR_{CT} (>0.8) revealed reduced probabilities of stenosis to 2 % and 14 %. Similarly at vessel-level (Fig. 3b), the probabilities of disease were increased or reduced by about 30 %.

4. Discussion

Our review has found that ML-FFR_{CT} owes a moderate discrimination capability to diagnosis coronary ischemia. The pooled sensitivity and specificity of the ML-FFR_{CT} at patient-level was 0.86 [95 % CI: 0.79, 0.91] and 0.87 [95 % CI: 0.76, 0.94], respectively, with pooled AUC 0.92 [95 % CI: 0.89–0.94]. At vessel-level, the pooled sensitivity and specificity was 0.80 [95 % CI: 0.74–0.84] and 0.84 [95 % CI: 0.77–0.89), respectively, with pooled AUC 0.88 [95 % CI: 0.85–0.91]. This shows that diagnosis at patient-level performs slightly better than that at

Table1
Study and Patient characteristics of the included studies.

Author	Year	Country/City	Study Design	Patient Characteristics (inclusion criteria)	Exclusion criteria	Patient Age, Years, Mean (SD)	Sample Size	No. Of Patients in analysis	Male [%]	AI method
Y Xue [15]	2022	China	retrospective	suspected or known stable CAD	non diagnostic CCTA images, FFR_{CT} calculation failure, previous revasularisation	61.9 (± 9.0)	522	484	71.5	cFFR v3.2 Siemens
H Yan [21]	2022	China	retrospective	at least one coronary artery with stenosis	poor image quality, previous revascularisation	56.6 (± 9.1)	1747	152	75.7	cFFR v3.0 Siemens
Y Li [16]	2022	China	retrospective	suspected CAD	previous revasularisation	57.3 (± 8.8)	73	73	72.6	multilayer perceptron network (MLP) and biditectional multilayer recursive neural network (BRNN)
HJ Koo [22]	2021	South Korea	retrospective	suspected significant CAD	previous revasularisation	62.8 (± 9.6)	471	471	77.7	cFFR v2.1 Siemens
F Zhou [23]	2019	China	retrospective	known or suspected CAD	FFRCT calculation failure, previous revasularisation	61.2 (± 9.1)	284	104	72.1	cFFR v3.2 Siemens
Y Li [16]	2019	China	prospective	intermidiate-high pretest probability of obstructive CAD	previous revasularisation	67.0 (± 12.0)	139	72	73	cFFR v3.0 Siemens
J de Geer [24]	2019	South Korea, the Netherlands, Sweden, Poland, USA	retrospective	underwent both clinical CCTA and ICA from MACHINE consortium database	non diagnostic CCTA images, FFR_{CT} calculation failure, failure to compute $cFFR_{ML}$	63.0 (range 56–69)	417	351	NA	cFFR v2.1 Siemens
A Kurata [17]	2019	Japan	retrospective	calcium score < 1500	poor image quality, previous revascularisation, Agatson score > 1500	70.2 (± 10.3)	206	74	76	cFFR v3.0 Siemens
PL von Knebel Doeberitz [25]	2019	Germany	retrospective	suspected or known stable angina patients	non diagnostic CCTA images, previous revasularisation	61.0 (± 10.0)	156	84	65	cFFR v2.1 Siemens
X Hu [26]	2018	China	retrospective	undergone CCTA and iFFR	non diagnostic CCTA images, >30 % stenosis in left main coronary artey, >80 % stenosi in FFR-measured coronoary ateries, previous revasularisation	62.0 (± 18.0)	156	105	69.5	cFFR v3.0 Siemens
C Tesche [18]	2018	Germany	retrospective	suspected CAD	non diagnostic CCTA images, previous revasularisation, severely reduced left ventricular function	62.0 (± 11.0)	166	85	62	cFFR v1.4 Siemens
A Coenen [27]	2018	South Korea, the Netherlands, Sweden, Poland, USA	mixed		non diagnostic CCTA images, FFR_{CT} calculation failure, failure to compute $cFFR_{ML}$	62.3 (± 9.3)	417	351	74	cFFR v2.1 Siemens
C Tesche [19]	2017	Germany	retrospective	suspected CAD	non diagnostic CCTA images, previous revasularisation	62.2 (± 10.6)	97	74	62	cFFR v2.1 Siemens
PP Xu [28]	2019	China	retrospective	suspected or known stable CAD	previous revasularisation, unreliable iFFR and software abilities	61.0 (± 5.5)	463	437	71.2	cFFR v3.2.0 Siemens
KK Kumamaru [14]	2022	Japan	retrospective	underwent CCTA	previous revasularisation	68.4 (± 10.5)	1052	131	71	fully connected neural netwrok
LM Tang [30]	2022	China	retrospective	known CAD	poor image quality, previous revascularisation	62 (±5.4)	144	144	66.7	DeepVessel, Keya Medical
ZQ Wang [20]	2019	China	prospective	undergone CCTA and iFFR	non adequate ICA and FFR measurements, previous revasularisation	68.8 (± 8.6)	68	63	50.8	DeepVessel, Keya Medical

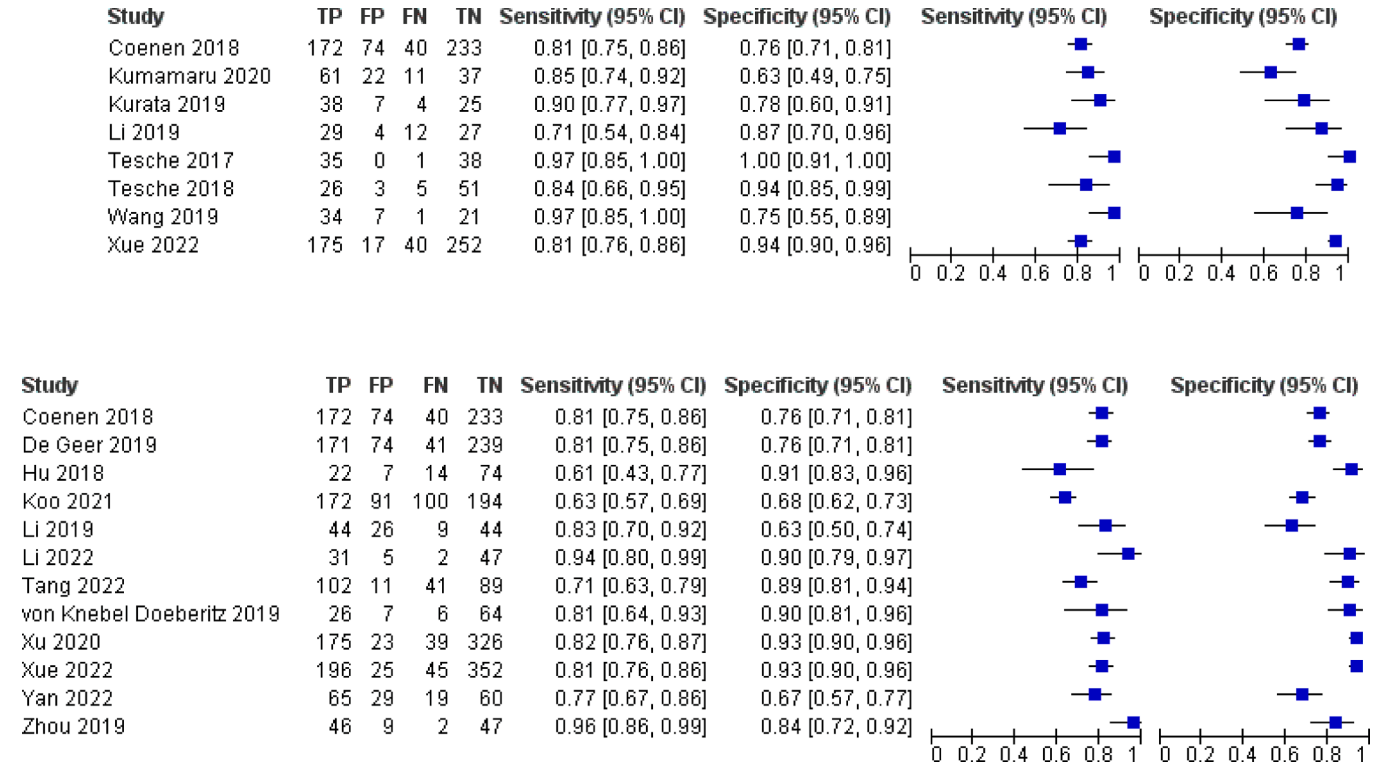

Fig. 1. *Forest plot of ML-FFR$_{CT}$. CI = confidence interval; FN = false negative; FP = false positive; TN = true negative; TP = true positive.*

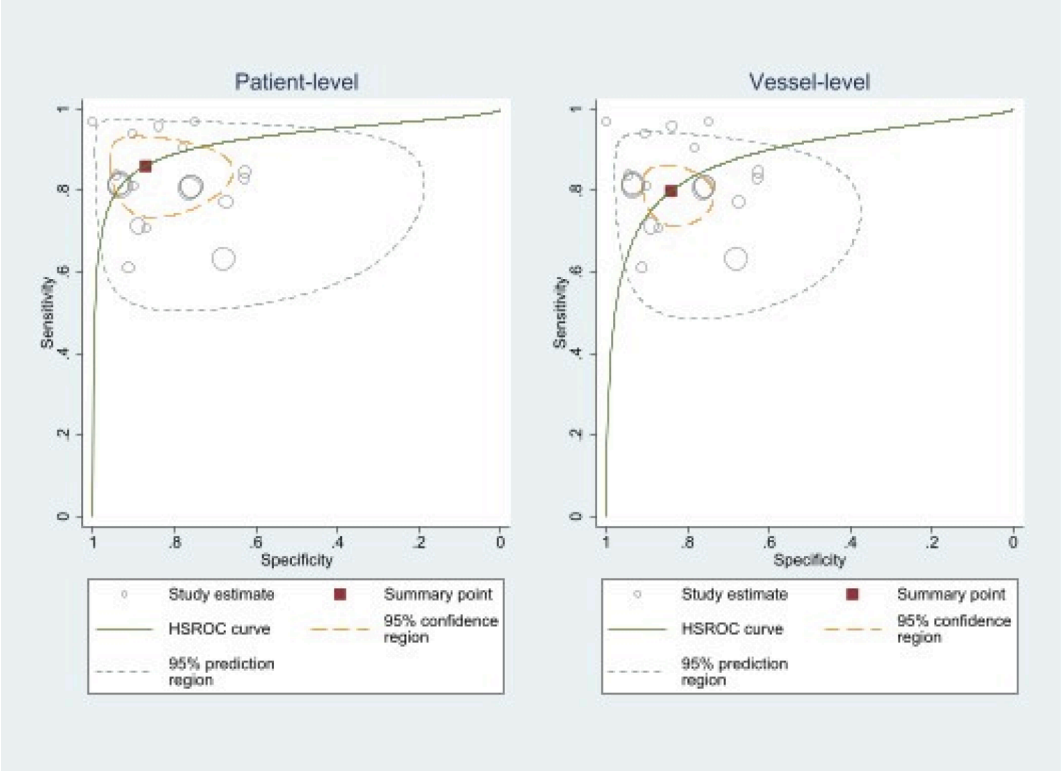

Fig. 2. *SROC evaluating ML-FFR$_{CT}$. SROC = summary receiver operating characteristic.*

vessel-level. This could be explained by the fact that a patient can have more than one affected vessel and the worst vessel only is considered when determining the severity of the stenosis for that patient. Looking at patient-level data would yield a higher diagnostic accuracy.

It is clear, that non-invasive procedures have the preference over invasive ones. CCTA is a non-invasive method to for the detection of anatomically significant CAD. But while this method has a high sensitivity, it has a low specificity, which results in more unnecessary

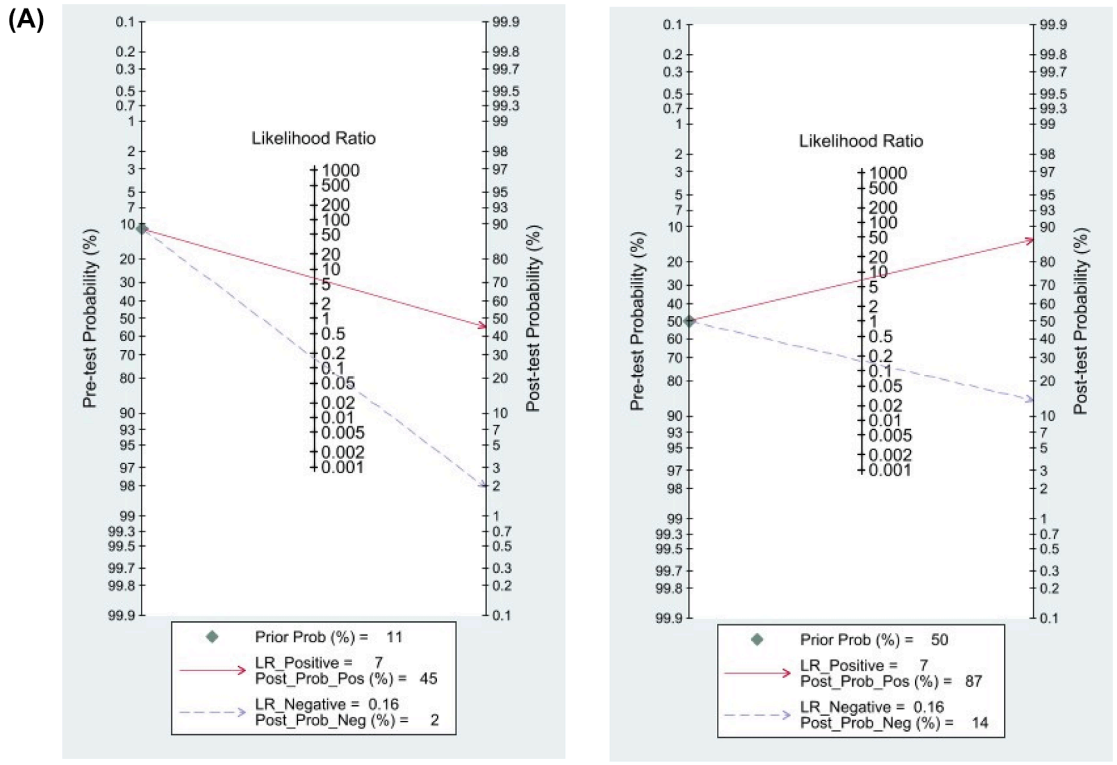

Fig. 3a. *Fagan's plot of ML-FFR$_{CT}$ (patient-level).*

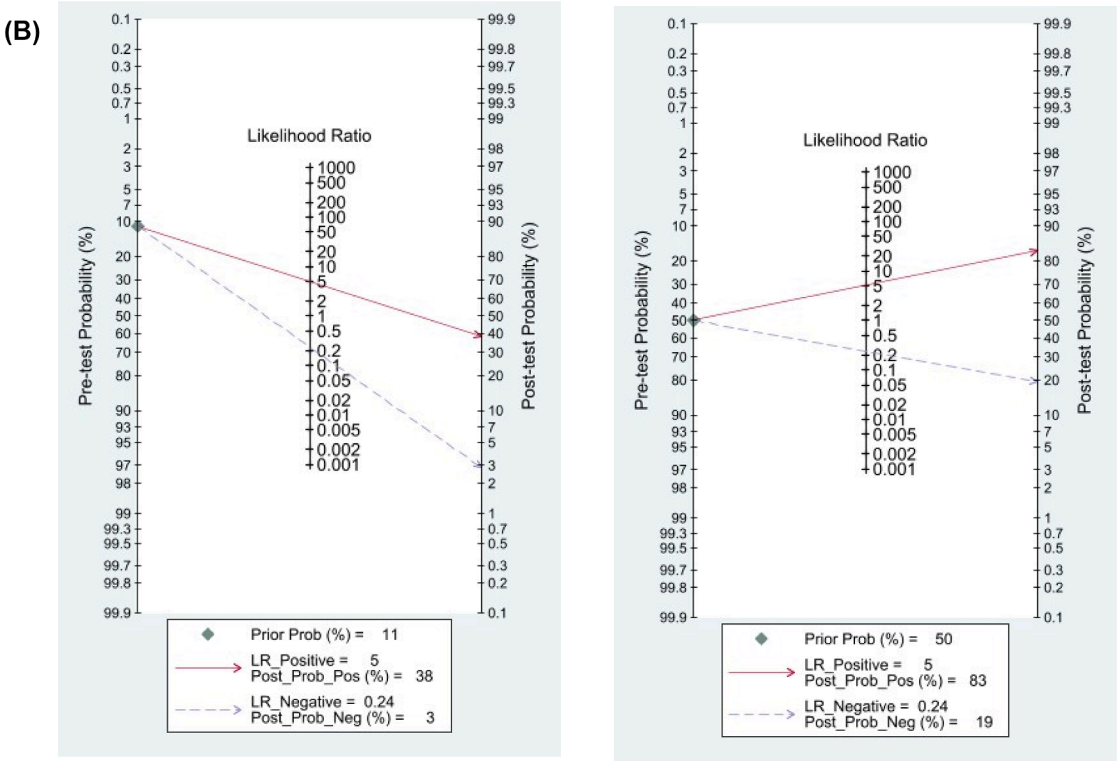

Fig. 3b. *Fagan's plot of ML-FFR$_{CT}$ (vessel-level). LR = likelihood ratio.*

revascularisation [32]. ML-FFR$_{CT}$, however, has shown higher specificity while maintaining the same level of sensitivity, which reduces false positive cases (false positive rate: 27/717 (3.8 %) and 105/1745 (6.0 %), per patient and per vessel respectively). As misdiagnosing patients without coronary ischemia as false positive will lead to unnecessary invasive procedure that may cause unnecessary cost and harm. These parameters can differ, as all studies were not performed under the same circumstances, using different ML-algorithms and

different radiologist interpreting the CCTA results. But it does indicate a trend.

The area under the ROC (AUC) shows the ML-FFR_{CT} has a very high discriminative power (0.92) and could reliably differentiate patients at patient level, and at vessel level also indicates good power (0.88). In addition, both positive and negative likelihood ratio have good performance and show moderate effect of ML-FFR_{CT} on improving diagnostic certainty. All the evidence has showed that ML-FFR_{CT} is a viable alternative to iFFR, estimating the FFR value without the need of invasive measurements and the extra administration of medications or contrast agents. It also can cut cost in terms of the procedure and patient care. Also, in principle, ML-FFR_{CT} could be executed in real-time, providing the clinician with immediate results [33]. This is not the case with other non-invasive FFR methods, such as CFD. CFD typically takes up to 2 h to produce results, excluding the time for manual annotation of the lumen from the CCTA images. However, crucially, the accuracy of ML-FFR_{CT} does strongly depend on the choice of ML-algorithm, the image quality, and the size of the training-set. Furthermore, what in general is not considered with the ML-approach is the amount of time and data needed to train the network.

One disconcerting reading of the data presented in this analysis is the declining accuracy reported in the studies over time, especially when determining the diagnostic accuracy for the detection of hemodynamic relevance of stenosis by CCTA, while the sensitivity stays the same over the period 2016–2022 (Supplementary Figure S3). The accuracy of ML-FFR_{CT} seems to increase slightly over the years for the patient and vessel-level diagnosis, apart from 2 studies which reported an accuracy of 63.0 % and 63.1 % per vessel and using the Siemens ML-algorithm, version 3.0 and 2.1, respectively [21,22]. However, it seems the accuracy of ML-FFR_{CT} at patient-level hits a ceiling at around 90 %, with no progress over the years from 2019 (Supplementary Figure S4a). The same can be said when looking at the vessel-level diagnosis accuracy (Supplementary Figure S4b).

There is also a big swing in accuracy, which could be caused by the inclusion or exclusion criteria for the respective studies. The difference between the studies by Xue *et al.* [15] and Yan *et al.* [21] is the image quality, while studies by Xue *et al.* [15] and Koo *et al.* [22] have nearly the same number of patients included. In the study by Xue *et al.* [15] the patient is excluded when there is a non-diagnostic CCTA image, while for the study by Yan *et al.* [21] the patient is excluded when the quality of CT image was too poor to extract the coronary artery tree for FFR_{CT} (Supplementary SA). There is a subtle difference between these two exclusion criteria as the more "poor images" removed from the dataset, the higher the accuracy will be. Also, several studies mentioned excluded patients of whom the ML-FFR_{CT} could not be calculated [15,17,24,27]. This will also affect the accuracy of diagnosis, for the better.

Although this review showed an overall good performance of ML application on diagnosis of stenosis, there are, however, some limitations to consider when using ML. For one, image quality is considered when selecting the study population, which means AI will have a higher accuracy then when compared to iFFR, as the clinician has to determine the level of stenosis for every patient. A reduction in image quality by various factors, such as noise, motion, beam hardening, will affect the outcome of the ML-algorithm. On the other hand, the quality of training data will affect the quality of algorithm per se. Some of the studies only had a relatively small subset for training data (*e.g.* study by Kumamaru *et al.* [14] in our analysis). The Siemens ML-algorithm is already pre-trained, but with synthetic generated coronary geometries.

Also, physiological and protocol-dependent factors, such as heart rate control, blood pressure, contrast enhancement methods may affect the FFR_{CT} value and therefore the diagnosis of the patient [34]. A threshold of 50 % of stenosis tends to lead to high sensitivity but low specificity for hemodynamically significant CAD [19]. Maybe it would be better to consider a higher threshold of DS. 70 % stenosis tends to have a far better specificity than a 50 % stenosis.

5. Limitations

Although the sensitivity analysis showed robust results on diagnostic parameters, the heterogeneity remains high. The high level of variation between studies may be due to difference in study design, patient profile, operation of measurement, *etc*. RoB level might be one of the factors in heterogeneity at patient level studies as by removing high RoB studies, the heterogeneity decreased and positive likelihood ratio increased. Significant heterogeneity was also observed in other systematic reviews that may indicate the high variations among studies on stenosis diagnosis [35,36].

Patients in 2 studies by Li 2022 *et al.* [16] and de Geer *et al.* [24] were selected on whether they fall below or above the 50 % stenosis cut off. That means that groups of patients are excluded for visual inspection of the CCTA results. Although the purpose of an ML-algorithm would be to perform at all levels of stenosis.

There would have been a higher inclusion number of studies, if authors of these studies would have included information about sensitivity, specificity and predictive values when describing the diagnostics result [37]. With more evidence from the studies on use of AI in stenosis diagnosis will enable a more precise estimation of the true effect. On the other hand, among the total 146 authors contributed to the 17 selected studies, a few authors have dominated the field (Bayer RR 2nd, de Cecco CN, Schoepf UF, Tesche C, Varga-Szemes A and Yu MM). In particular, 2 authors (Schoepf and Tesche) have contributed 6 of the 17 (35 %) publications.

6. Conclusion

Although heterogeneity exists, the presented data shows that non-invasive ML-FFR_{CT} has a high diagnostic accuracy for patients with known or suspected CAD, comparable to iFFR. As such, it can serve as a triage tool that can complement clinical history, symptom score, to determine the presence and severity of underlying stenosis, to avoid unnecessary invasive procedures for the patients.

Still more needs to be considered to increase the score of true positive when using ML-algorithm. An improved true positive score will result in an even lower cases of revascularisation. One way to achieve this is by combining CCTA and ML results, as highlighted in Han *et al* [38].

7. Authors' statement

All authors take responsibility for all aspects of the reliability and freedom from bias of the data presented and their discussed interpretation.

Author contributions

DvN selected the studies, extracted and formatted the data, did part of the data analysis and wrote the manuscript; GL gave advise on writing *meta*-analysis, did data analysis and wrote the sections on data reliability and bias; LS was second reviewer for the study selection and edited the manuscript. LuS, RST, LT, MSY, LB, PC, FK, MC, TC, SYT, ZL contributed to the concept of the study and manuscript editing. ZL was also the third reviewer for the study selection.

Funding

This study is supported by Singapore National Medical Research Council (NMRC) (MOH-000358).

CRediT authorship contribution statement

Danny van Noort: Writing – review & editing, Writing – original draft, Visualization, Validation, Investigation, Formal analysis, Data curation, Conceptualization. **Liang Guo:** Writing – review & editing,

Supervision, Project administration, Funding acquisition, Conceptualization. **Shuang Leng:** Writing – review & editing, Validation. **Luming Shi:** Writing – review & editing, Supervision, Funding acquisition. **Ru-San Tan:** Writing – review & editing, Funding acquisition. **Lynette Teo:** Writing – review & editing, Funding acquisition. **Min Sen Yew:** Writing – review & editing, Funding acquisition. **Lohendran Baskaran:** Writing – review & editing, Supervision, Funding acquisition. **Ping Chai:** Writing – review & editing, Funding acquisition. **Felix Keng:** Writing – review & editing, Funding acquisition. **Mark Chan:** Writing – review & editing, Funding acquisition. **Terrance Chua:** Writing – review & editing, Funding acquisition. **Swee Yaw Tan:** Writing – review & editing, Funding acquisition. **Liang Zhong:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijcha.2024.101528.

References

[1] D. Mozaffarian, E.J. Benjamin, A.S. Go, et al., Heart disease and stroke statistics-2016 update: A report from the American heart association, Circulation 133 (2016) e38iseas.

[2] N.H.J. Pijls, B. de Bruyne, M. El Gamal, et al., Fractional flow reserve: the ideal parameter for evaluation of coronary, myocardial, and collateral blood flow by pressure measurements at PTCA, J. Interv Cardiol 6 (4) (1993) 331–344.

[3] M.J. Kern, A. Lerman, J.-W. Bech, et al., Physiological assessment of coronary artery disease in the cardiac catheterization laboratory: a scientific statement from the American heart association committee on diagnostic and interventional cardiac catheterization, **Council on Clinical Cardiology.** *Circulation* 114 (2006) 1321–1341.

[4] B. de Bruyne, N.H.J. Pijls, B. Kalesan, et al., Fractional flow reserve-guided PCI versus medical therapy in stable coronary disease, N. Engl J. Med. 367 (2012) 991–1001.

[5] T.F. Members, G. Montalescot, U. Sechtem, et al., ESC guidelines on the management of stable coronary artery disease: the Task Force on the management of stable coronary artery disease of the European Society of Cardiology, Eur. Heart J. 2013 (34) (2013) 2949–3003.

[6] M. Tavakol, S. Ashraf, S.J. Brener, Risks and complications of coronary angiography: a comprehensive review, Glob. J. Health Sci. 4 (1) (2012) 65–93.

[7] M.C. Williams, J.P. Earls, H. Hecht, Quantitative assessment of atherosclerotic plaque, recent progress and current limitations, J. Cardiovasc Comput. Tomogr 16 (2) (2022) 124–137.

[8] A. Arbab-Zadeh, J. Hoe, Quantification of coronary arterial stenosis by multidetector CTangiography in comparison with conconventional angiography methods, caveats, and implications, JACC Cardiovasc Imaging 4 (2011) 191–202.

[9] M. Andrew, H. John, The challenge of coronary calcium on coronary computed tomographic angiography (CCTA) scans: effect on interpretation and possible solutions, Int. J. Cardiovasc Imaging 31 (2015) 145–157.

[10] C.A. Taylor, T.A. Fonte, J.K. Min, Computational fluid dynamics applied to cardiac computed tomography for noninvasive quantification of fractional flow reserve, J. Am. Coll Cardiol 61 (2013) 2233–2241.

[11] J.M. Zhang, D. Shuang, L. Baskaran, et al., Advanced analyses of computed tomography coronary angiography can help discriminate ischemic lesions, Int. J. Cardiol 267 (2018) 208–214.

[12] J.M. Zhang, L. Zhong, T. Luo, et al., Simplified models of non-invasive fractional flow reserve based on CT images, PLOS ONE 11 (5) (2016) e0153070.

[13] L. Itu, S. Rapaka, T. Passerini, et al., A machine-learning approach for computation of fractional flow reserve from coronary computed tomography, J. Appl. Physiol. 121 (2016) 42–52.

[14] K.K. Kumamaru, S. Fujimoto, Y. Otsuka, et al., Diagnostic accuracy of 3D deep-learning-based fully automated estimation of patient-level minimum fractional flow reserve from coronary computed tomography angiography, Eur. Heart J. Cardiovasc Imaging 21 (4) (2020) 437–445.

[15] Y. Xue, M.W. Zheng, Y. Hou, et al., Influence of diabetes mellitus on the diagnostic performance of machine learning-based coronary CT angiography-derived fractional flow reserve: a multicenter study, Eur. Radiol. 32 (6) (2022) 3778–3789.

[16] Y. Li, M. Yu, X. Dai, et al., Detection of hemodynamically significant coronary stenosis: CT myocardial perfusion versus machine learning CT fractional flow reserve, Radiology 293 (2) (2019) 305–314.

[17] A. Kurata, N. Fukuyama, K. Hirai, et al., On-site computed tomography-derived fractional flow reserve using a machine-learning algorithm - clinical effectiveness in a retrospective multicenter cohort, Circ. J. 83 (7) (2019) 1563–1571.

[18] C. Tesche, C.N. De Cecco, S. Baumann, et al., Coronary CT angiography-derived fractional flow reserve: Machine learning algorithm versus computational fluid dynamics modeling, Radiology 288 (1) (2018) 64–72.

[19] C. Tesche, R. Vliegenthart, T.M. Duguay, et al., Coronary computed tomographic angiography-derived fractional flow reserve for therapeutic decision making, Am. J. Cardiol 120 (12) (2017) 2121–2127.

[20] Z.Q. Wang, Y.J. Zhou, Y.X. Zhao, et al., Diagnostic accuracy of a deep learning approach to calculate FFR from coronary CT angiography, J. Geriatr Cardiol 16 (1) (2019) 42–48.

[21] H. Yan, Y. Gao, N. Zhao, et al., Change in computed tomography-derived fractional flow reserve across the lesion improve the diagnostic performance of functional coronary stenosis, Front Cardiovasc Med. 8 (2022) 788703.

[22] H.J. Koo, J.W. Kang, S.J. Kang, et al., Impact of coronary calcium score and lesion characteristics on the diagnostic performance of machine-learning-based computed tomography-derived fractional flow reserve, Eur. Heart J. Cardiovasc Imaging 22 (9) (2021) 998–1006.

[23] F. Zhou, Y.N. Wang, U.J. Schoepf, et al., Diagnostic performance of machine learning based CT-FFR in detecting ischemia in myocardial bridging and concomitant proximal atherosclerotic disease, Can. J. Cardiol 35 (11) (2019) 1523–1533.

[24] J. De Geer, A. Coenen, Y.H. Kim, et al., Effect of tube voltage on diagnostic performance of fractional flow reserve derived from coronary CT angiography with machine learning: Results from the MACHINE registry, AJR Am. J. Roentgenol 213 (2) (2019) 325–331.

[25] P.L. von Knebel Doeberitz, C.N. De Cecco, U.J. Schoepf, et al., Coronary CT angiography-derived plaque quantification with artificial intelligence CT fractional flow reserve for the identification of lesion-specific ischemia, Eur. Radiol 29 (5) (2019) 2378–2387.

[26] X. Hu, M. Yang, L. Han, Y. Du, Diagnostic performance of machine-learning based computed fractional flow reserve (FFR) derived from coronary computed tomography angiography for the assessment of myocardial ischemia verified by invasive FFR, Int. J. Cardiovasc Imaging 34 (12) (2018) 1987–1996.

[27] A. Coenen, Y.H. Kim, M. Kruk, et al., Diagnostic accuracy of a machine-learning approach to coronary computed tomographic angiography-based fractional flow reserve: Result from the MACHINE consortium, Circ Cardiovasc Imaging 11 (6) (2018) e007217.

[28] P.P. Xu, J.H. Li, F. Zhou, et al., The influence of image quality on diagnostic performance of a machine learning-based fractional flow reserve derived from coronary CT angiography, Eur. Radiol 30 (5) (2020) 2525–2534.

[29] Y. Li, H. Qiu, Z. Hou, et al., Additional value of deep learning computed tomographic angiography-based fractional flow reserve in detecting coronary stenosis and predicting outcomes, Acta Radiol 63 (1) (2022) 133–140.

[30] L.M. Tang, F. Liu, T.Y. Dong, F. Yang, S.J. Cui, Synergistic value of fractional flow reserve and low–density non–calcified plaque based on coronary computed tomography angiography for the identification of lesion–specific ischemia, Exp. Ther. Med. 24 (5) (2022) 701.

[31] Centers for Disease Control and Prevention and National Center for Health Statistics. **Summary Health Statistics: National Health Interview Survey, 2018: table A-1**. Accessed March 22, 2023.

[32] Y.B. Song, A. Arbab-Zadeh, M.B. Matheson, et al., Contemporary discrepancies of stenosis assessment by computed tomography and invasive coronary angiography. Analysis of the CORE320 international study, Circ Cardiovasc Imaging 12 (2019) e007720.

[33] J. Liao, L. Huang, M. Qu, B. Chen, G. Wang, Artificial intelligence in coronary CT angiography: current status and future prospects, Front Cardiovasc Med. 9 (2022) 896366.

[34] J. Leipsic, T.H. Yang, A. Thompson, et al., CT angiography (CTA) and diagnostic performance of noninvasive fractional flow reserve: results from the determination of fractional flow reserve by anatomic CTA (DeFACTO) Study, AJR Am. J. Roentgenol 202 (2014) 989–994.

[35] W. Man, J. Hu, Z. Zhao, et al., Diagnostic performance of instantaneous wave-free ratio for the evaluation of coronary stenosis severity confirmed by fractional flow reserve: A PRISMA-compliant meta-analysis of randomized studies, Medicine 95 (36) (2016) e4774.

[36] W. Zuo, M. Yang, Y. Chen, A. Xie, L. Chen, G. Ma, Meta-analysis of diagnostic performance of instantaneous wave-free ratio versus quantitative flow ratio for detecting the functional significance of coronary stenosis, Biomed Res. Int. 18 (2019) 5828931.

[37] R. Trevethan, Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice, Front Public Health 5 (2017) 307.

[38] D. Han, J.H. Lee, A. Rizvi, et al., Incremental role of resting myocardial computed tomography perfusion for predicting physiologically significant coronary artery disease: A machine learning approach, J. Nucl. Cardiol 25 (1) (2018) 223–233.