

<https://doi.org/10.1038/s41746-026-02338-6>

Multicenter evaluation of interpretable AI for coronary artery disease diagnosis from PET biomarkers

Check for updates

Wenhao Zhang^{1,2}, Jacek Kwiecinski³, Aakash Shanbhag^{1,4}, Robert J. H. Miller^{1,5}, Shiva Mostafavi¹, Giselle Ramirez¹, Jirong Yi¹, Donghee Han¹, Damini Dey¹, Dominika Grodecka^{1,6}, Kajetan Grodecki¹, Mark Lemley¹, Paul Kavanagh¹, Joanna X. Liang¹, Jianhang Zhou¹, Valerie Builoff¹, Jon Hainer⁷, Sylvain Carre⁷, Leanne Barrett⁷, Andrew J. Einstein⁸, Stacey Knight⁹, Steve Mason⁹, Viet T. Le⁹, Wanda Acampa¹⁰, Samuel Wopperer¹¹, Panithaya Chareonthaitawee¹¹, Daniel S. Berman¹, Marcelo F. Di Carli^{7,12} & Piotr J. Slomka^{1,12}✉

Positron emission tomography (PET)/computed tomography (CT) for myocardial perfusion imaging (MPI) provides multiple imaging biomarkers, often evaluated separately. We developed an artificial intelligence (AI) model integrating key clinical PET MPI parameters to improve the diagnosis of obstructive coronary artery disease (CAD). From 17,348 patients undergoing cardiac PET/CT across four sites, 1664 with invasive coronary angiography and no prior CAD were retrospectively analyzed. Coronary artery calcium (CAC) scores were derived from CT attenuation correction maps, and XGBoost model was trained on one site using 10 image-derived parameters: CAC, stress/rest left ventricular ejection fraction, stress myocardial blood flow (MBF), myocardial flow reserve (MFR), ischemic and stress total perfusion deficit (TPD), transient ischemic dilation ratio, rate pressure product, and sex. External validation was performed across three independent sites. In the testing cohort ($n = 1278$; CAD prevalence 53%), the AI model achieved an area under the receiver operating characteristic curve (AUC) of 0.83 (95% CI: 0.81–0.85), outperforming experienced physicians (0.80, $p = 0.02$) and individual biomarkers such as ischemic TPD (0.79, $p < 0.001$) and MFR (0.75, $p < 0.001$). Performance was consistent across sex, body mass index, and age. AI integrating perfusion, flow, and CAC scoring improves PET MPI diagnostic accuracy, offering automated and interpretable predictions for CAD diagnosis.

Positron emission tomography (PET) myocardial perfusion imaging (MPI), especially when combined with computed tomography (CT), plays a pivotal role in assessing patients with suspected coronary artery disease (CAD). The wealth of data provided by cardiac PET/CT includes: (1) perfusion imaging, which depicts the extent of ischemia; (2) measurements of absolute

myocardial blood flow (MBF), quantifying the volume of blood flow per unit time per myocardial mass (typically mL/min/g) and enhancing diagnostic accuracy beyond perfusion imaging alone^{1–5}; (3) functional information; and (4) coronary artery calcium (CAC), a crucial indicator of atherosclerotic burden, all of which offer significant insights for risk stratification⁶. The

¹Artificial Intelligence in Medicine Research Center, Departments of Biomedical Sciences, Medicine, and Cardiology, Cedars-Sinai Medical Center, Los Angeles, CA, USA. ²School of Automation and Intelligent Sensing, Shanghai Jiao Tong University, Shanghai, China. ³Department of Interventional Cardiology and Angiology, Institute of Cardiology, Warsaw, Poland. ⁴Signal and Image Processing Institute, Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA, USA. ⁵Department of Cardiac Sciences, University of Calgary, Calgary, AB, Canada. ⁶Department of Anesthesiology and Intensive Care, Medical University of Warsaw, Warsaw, Poland. ⁷Department of Radiology, Division of Nuclear Medicine and Molecular Imaging, Brigham and Women's Hospital, Boston, MA, USA. ⁸Division of Cardiology, Department of Medicine, and Department of Radiology, Columbia University Irving Medical Center/New York-Presbyterian Hospital, New York, NY, USA. ⁹Intermountain Medical Center Heart Institute, Intermountain Healthcare, Murray, UT, USA. ¹⁰Department of Advanced Biomedical Sciences, University of Naples Federico II, Naples, Italy. ¹¹Department of Cardiovascular Medicine, Mayo Clinic, Rochester, MN, USA. ¹²These authors contributed equally: Marcelo F. Di Carli, Piotr J. Slomka. ✉e-mail: Piotr.Slomka@cshs.org

latter can be derived automatically from CT attenuation correction (CTAC) scans, which are routinely acquired as part of MPI exams on modern PET/CT systems^{3,7}.

However, currently, the assessment of CAD with PET/CT MPI does not optimally leverage the combined diagnostic power of these various imaging markers, including CAC^{3,4,8,9}. Recognizing this gap, we developed an artificial intelligence (AI) model that integrates a parsimonious set of 10 common PET MPI parameters—features often used intuitively but not systematically combined in clinical practice. While clinicians assess these parameters during interpretation, synthesizing all relevant data into a single interpretation can be challenging. In fact, there remains debate regarding the best methods to integrate just three features (CAC, relative perfusion, and blood flow)¹⁰, highlighting the need for an objective, automated approach.

Our AI approach harnesses deep learning and machine learning to comprehensively integrate and analyze key PET–CT parameters—including calcium burden, perfusion, MBF, and functional metrics—enabling a more robust assessment of CAD. We based our model on standard PET imaging measurements rather than raw images to ensure generalizability, interpretability, and transparency—critical factors for clinical adoption. Furthermore, we rigorously validated our model on a large external testing cohort from three sites to confirm its robustness. To our knowledge, this work represents the first multicenter, externally validated, AI-driven cardiac PET MPI analysis.

Results

Study population

Patient characteristics are shown in Table 1. Among the 1664 participants, 1085 (65%) were male, and the median age was 68 with an interquartile range (IQR) of (61, 75). Of these patients, 386 were allocated for model training and optimization, while 1278 from separate institutions were used for external testing. The training data were complete, containing no missing values, whereas the percentages of missingness in the testing set were $\leq 1\%$ (Supplementary Table 1). For univariate comparisons in the external testing set, missing values of total perfusion deficit (TPD) were imputed using the corresponding summed scores with scores/68*100%, and missing values of flows and CAC were imputed with their medians. The prevalence of CAD was significantly higher in the internal training set than in the external testing set (61% vs. 53%, $p = 0.004$).

Myocardial imaging analysis parameters

In both the internal training and external testing sets, patients with CAD demonstrated significantly higher median ischemic TPD, stress TPD, and transient ischemic dilation ratio (TID) compared to those without CAD ($p < 0.001$ for all in the testing set; $p < 0.001$ for TPDs and $p = 0.022$ for TID in the training set) (Table 2). Additionally, patients with CAD had notably lower stress MBF ($p < 0.001$), myocardial flow reserve (MFR) ($p < 0.001$), and median stress ejection fraction ($p = 0.011$ in training, $p < 0.001$ in testing).

Coronary artery calcium

Of the internal training cohort ($n = 386$), CAC was 0 in 58 (15.0%) patients, >0 –100 in 58 (15.0%), >100 –400 in 79 (20.5%), and $CAC > 400$ in 191 (49.5%) subjects. In the external testing set ($n = 1278$), 288 (22.5%) had a CAC of 0, 231 (18.1%) had a CAC between >0 –100, 264 (20.7%) had a CAC between >100 –400, and 495 (38.7%) had a CAC > 400 (Table 1).

Model performance

In external testing, we compared the AI model with clinical measurements (Fig. 1). The comprehensive machine learning model (area under the receiver operating characteristic curve [AUC] 0.83, [0.81–0.85]) outperformed clinical score from experienced physicians (0.80 [0.77–0.82], $p = 0.02$), ischemic TPD (0.79, [0.76–0.81], $p < 0.001$), MFR (0.75, [0.72–0.78], $p < 0.001$), stress MBF (0.75, [0.72–0.77], $p < 0.001$), and CAC (0.69, [0.66–0.72], $p < 0.001$) (Fig. 2a, Supplementary Table 2). Figure 2b shows that the AI model achieved significantly higher sensitivity ($p < 0.001$)

than quantitative thresholds, including a minimum 3-vessel MFR < 2 , stress MBF < 1.8 , and ischemic TPD $\geq 5\%$, when matched for specificity in the external testing set. The AI thresholds and diagnostic accuracies were reported in Supplementary Table 3. This improvement was driven by the model's ability to identify more high-risk patients classified by Duke 6 and Duke 5 criteria, compared to quantitative measurements at matched specificity (Supplementary Table 4). Supplementary Table 5 reports the performance assessment stratified by site and isotope, further highlighting the generalizability of the model. This is likely because we utilized tracer-specific normal limits for perfusion quantification and dedicated kinetic models for flow quantification, which ensured compatible quantitative values across the two tracers. We also compared the Extreme Gradient Boosting (XGBoost) model with the conventional logistic regression model in Supplementary Table 6, obtaining similar results. The calibration performance was presented in Supplementary Fig. 1.

The top features driving the prediction were ischemic and stress TPD, CAC, and MFR using shapley additive explanations (SHAP) values in Supplementary Fig. 2. In a predefined subgroup analysis, the AI models exhibited comparable performance across various subgroups: female and male patients (AUC 0.83 vs. 0.83, $p = 0.65$), patients classified as non-obese (body mass index [BMI] < 30 kg/m²) and obese (BMI ≥ 30 kg/m²) (AUC 0.84 vs. 0.81, $p = 0.10$), and older (age ≥ 65 years) and younger (age < 65 years) (AUC 0.84 vs. 0.81, $p = 0.36$) (Supplementary Fig. 3). Additionally, the AI model was evaluated for its performance among patients not referred to 180-day invasive coronary angiography (ICA), with outcomes assessed against major adverse cardiovascular events up to 5 years (Supplementary Fig. 4).

Individualized Inference by the AI model

In Fig. 3, two cases illustrate the ability of the AI model to integrate multiple PET MPI parameters for CAD diagnosis. Figure 3a demonstrates how the model accurately identifies obstructive CAD, emphasizing MFR as a key driver. Figure 3b highlights the capability of the model to correctly rule out CAD despite borderline TPD and clinician-assigned stress scores. These examples demonstrate the model's effectiveness in leveraging integrated biomarkers to align predictions with patient outcomes.

Discussion

PET/CT, with its unique ability to quantify absolute MBF, allows robust detection of obstructive coronary stenoses across the spectrum of CAD, including diffuse atherosclerosis, microvascular disease, or multivessel disease. Consequently, PET is the fastest-growing non-invasive cardiac imaging test¹¹. However, to take full advantage of PET/CT, it is necessary to combine complex data, which is currently not performed systematically.

In this study, we developed, tested, and externally validated an interpretable AI model that integrates key quantitative PET MPI parameters—including stress/rest perfusion, myocardial flow, ejection fraction, AI-derived calcium scores, and sex—into a unified assessment. In multicenter external validation, our AI approach outperformed both expert physicians (provided with complete imaging/clinical data during assessment) and conventional quantitative methods. The AI model demonstrated improved sensitivity and superior detection of high-risk patients, highlighting its strong generalizability despite inter-center variability. To avoid the perception of AI as a 'black box', we have provided both patient-level and model-level explanations by highlighting the most influential factors in predicting significant CAD, enhancing physician understanding and confidence in AI-driven predictions. Collectively, these characteristics underscore the tremendous potential of our AI model for a real-world, practical approach to diagnosis, incorporating all major factors currently considered by physicians, including calcium derived by deep learning from CT attenuation maps.

To date, several single-site studies have investigated the benefit of integrating subsets of PET/CT MPI parameters^{3,4,9,12–16}. Initial studies focused on optimizing risk stratification by combining MBF and perfusion data^{3,12,16}. Gupta et al. established that MFR serves as a stronger predictor of

Table 1 | Patient characteristics

	All patients, <i>n</i> (%)	Training dataset, <i>n</i> (%)	External testing dataset, <i>n</i> (%)	<i>P</i> -value
<i>N</i> (%)	1664	386 (23.2%)	1278 (76.8%)	
Age, years	68 (61, 75)	71 (65, 80)	67 (60, 74)	<0.001
Male	1085 (65.2%)	250 (64.8%)	835 (65.3%)	0.9
Race				<0.001
American Indian or Alaska Native	11 (0.7%)	1 (0.3%)	10 (0.8%)	
Asian	36 (2.2%)	16 (4.1%)	20 (1.6%)	
Black or African American	102 (6.1%)	59 (15.3%)	43 (3.3%)	
Native Hawaiian or Other Pacific Islander	22 (1.3%)	2 (0.5%)	20 (1.6%)	
White	1410 (84.7%)	297 (76.9%)	1113 (87.1%)	
Unavailable	83 (5.0%)	11 (2.8%)	72 (5.6%)	
Current Smoking	312 (18.8%)	38 (9.8%)	274 (21.4%)	<0.001
BMI, kg/m ²	30 (26, 36)	28 (24, 32)	31 (26, 37)	<0.001
<i>Past medical history and comorbidities</i>				
Hypertension	1362 (81.9%)	298 (77.2%)	1064 (83.3%)	0.002
Diabetes mellitus	671 (40.3%)	144 (37.3%)	527 (41.2%)	0.2
Dyslipidemia	1198 (72.0%)	243 (63.0%)	955 (74.7%)	<0.001
Family history of CAD	561 (33.7%)	58 (15.0%)	503 (39.4%)	<0.001
Peripheral vascular disease	321(19.3%)	39 (10.1%)	282 (22.1%)	<0.001
Past myocardial infarction	0 (0%)	0 (0%)	0 (0%)	NA
Past PCI	0 (0%)	0 (0%)	0 (0%)	NA
CABG	0 (0%)	0 (0%)	0 (0%)	NA
<i>PET-CT acquisition parameters</i>				
Pharmacological stress agent				<0.001
Dipyridamole	19 (1.1%)	1 (0.3%)	18 (1.4%)	
Adenosine	131 (7.9%)	42 (10.9%)	89 (7.0%)	
Regadenoson	1489 (89.5%)	343 (88.9%)	1146 (89.7%)	
Dobutamine	25 (1.5%)	0 (0.0%)	25 (2.0%)	
Isotope				<0.001
¹³ N-ammonia	535 (32.2%)	0 (0.0%)	535 (41.9%)	
⁸² Rubidium	1,129 (67.8%)	386 (100.0%)	743 (58.1%)	
Tracer activities (stress dose, mCi)				<0.001
¹³ N-ammonia	18.0 (13.0, 23.0)	NA	18.0 (13.0, 23.0)	
⁸² Rubidium	35.3 (27.7, 42.9)	31.0 (23.0, 39.0)	37.5 (29.0, 46.0)	
Tracer activities (rest dose, mCi)				<0.001
¹³ N-ammonia	17.5 (13.0, 22.0)	NA	18.0 (13.0, 23.0)	
⁸² Rubidium	35.3 (27.7, 42.9)	31.0 (23.0, 39.0)	37.5 (29.0, 46.0)	
<i>PET-CT quantitative image analysis parameters</i>				
Stress MBF, mL/g/min	1.16 (0.77, 1.60)	1.27 (0.86, 1.79)	1.12 (0.75, 1.56)	<0.001
MFR	1.44 (1.03, 1.94)	1.30 (0.89, 1.85)	1.49 (1.08, 1.97)	<0.001
Stress TPD	9.9 (4.9, 18.3)	8.9 (4.2, 16.8)	10.4 (5.0, 18.9)	0.012
Stress TPD < 5%	434 (26.1%)	118 (30.6%)	316 (24.7%)	0.026
Ischemic TPD	6.5 (3.4, 11.8)	6.9 (3.3, 12.6)	6.3 (3.4, 11.4)	0.3
TID	1.05 (0.98, 1.13)	1.05 (0.99, 1.16)	1.05 (0.98, 1.12)	0.095
Stress EF, %	62 (47, 72)	63 (47, 72)	62 (47, 72)	0.6
Rest EF, %	61 (47, 70)	62 (47, 70)	61 (47, 70)	0.6
Coronary artery calcium score				<0.001
=0	346 (20.8%)	58 (15.0%)	288 (22.5%)	
>0–100	289 (17.4%)	58 (15.0%)	231 (18.1%)	
>100–400	343 (20.6%)	79 (20.5%)	264 (20.7%)	
>400	686 (41.2%)	191 (49.5%)	495 (38.7%)	
<i>Outcome</i>				
Presence of obstructive CAD	910 (54.7%)	236 (61.1%)	674 (52.7%)	0.004

Categorical values were expressed as *n* (%). Continuous values were expressed as mean ± standard deviation or median (interquartile range), based on the Shapiro–Wilk test for normality. Variables were compared using a Pearson χ^2 statistic for categorical variables and Mann–Whitney–Wilcoxon test for continuous variables.

MBF myocardial blood flow, *MFR* myocardial flow reserve, *NA* not applicable, *PCI* percutaneous coronary intervention, *TPD* total perfusion deficit, *TID* transient ischemic dilation.

cardiac mortality than maximal MBF in patients with stable CAD, with integrated assessment of MFR and MBF identifying distinct prognostic risk profiles, particularly highlighting MFR impairment as a key mortality indicator⁹. Gould et al. proposed the coronary flow capacity measure, which integrates regional stress MBF and MFR into one variable¹⁴. Using data from a single center, Singh et al. developed an explainable deep learning model for

all-cause mortality prediction, integrating polar maps of stress and rest perfusion, MBF, MFR, and spill-over fraction combined with cardiac volumes, singular indices, and sex¹⁵. These prior studies have consistently demonstrated that combining multiparametric PET data improves risk stratification, with more comprehensive models outperforming approaches relying on only two or three variables.

Table 2 | Parameters included in the AI model from the training set (n = 386) and external testing set (n = 1278)

	Overall, n (%)	CAD, n (%)	No CAD, n (%)	P-value
<i>Training set (n = 386)</i>				
Ischemic TPD	7 (3, 13)	10 (6, 16)	4 (2, 6)	<0.001
Stress TPD	9 (4, 17)	13 (7, 22)	5 (2, 9)	<0.001
Coronary calcium	372 (52, 1039)	651 (142, 1485)	150 (11, 572)	<0.001
Stress MBF	1.27 (0.86, 1.79)	1.07 (0.72, 1.47)	1.59 (1.20, 2.16)	<0.001
MFR	1.30 (0.89, 1.85)	1.07 (0.77, 1.56)	1.68 (1.28, 2.18)	<0.001
TID	1.05 (0.99, 1.16)	1.06 (1.00, 1.17)	1.04 (0.97, 1.12)	0.022
Rest EF, %	62 (47, 70)	62 (47, 69)	63 (46, 72)	0.4
RPP	9452 (8308, 11136)	9452 (8320, 11136)	9455 (8240, 11152)	>0.9
Stress EF, %	63 (47, 72)	62 (45, 71)	67 (49, 75)	0.011
Male	250 (65%)	156 (66%)	94 (63%)	0.6
<i>External testing set (n = 1278)</i>				
Ischemic TPD	6 (3, 11)	10 (6, 16)	4 (2, 6)	<0.001
Stress TPD	10 (5, 19)	16 (9, 25)	6 (3, 11)	<0.001
Coronary calcium	208 (10, 801)	423 (77, 1414)	73 (0, 426)	<0.001
Stress MBF	1.12 (0.75, 1.56)	0.88 (0.59, 1.26)	1.39 (1.03, 1.76)	<0.001
MFR	1.49 (1.08, 1.97)	1.22 (0.86, 1.61)	1.81 (1.39, 2.25)	<0.001
TID	1.05 (0.98, 1.12)	1.07 (1.01, 1.14)	1.03 (0.96, 1.10)	<0.001
Rest EF, %	61 (47, 70)	61 (48, 69)	61 (46, 72)	0.2
RPP	9268 (7811, 11079)	9292 (7853, 11097)	9221 (7786, 11054)	0.6
Stress EF, %	62 (47, 72)	60 (46, 70)	65 (49, 75)	<0.001
Male	835 (65%)	484 (72%)	351 (58%)	<0.001

Categorical values were expressed as n (%). Continuous values were expressed as mean ± standard deviation or median (interquartile range), based on the Shapiro–Wilk test for normality. Variables were compared using a Pearson χ^2 statistic for categorical variables and Mann–Whitney–Wilcoxon test for continuous variables. CAD coronary artery disease, EF ejection fraction, MBF myocardial blood flow, MFR myocardial flow reserve, RPP rate pressure product, TPD total perfusion deficit, TID transient ischemic dilation.

Fig. 1 | Central illustration. Artificial intelligence (AI) model integrating fully automated PET flow, perfusion quantification, and deep learning (DL)-based coronary artery disease in all patients undergoing myocardial perfusion imaging positron emission computed tomography/computed tomography (PET/CT). Receiver-operating characteristic curve for coronary artery disease (CAD) diagnosis and area under the receiver-operating characteristic curve values of the AI model vs. clinical score, ischemic total perfusion deficit (TPD), myocardial flow reserve (MFR), and coronary calcium in external validation assessment (n = 1278). Clinical score was defined as the summed difference score from the 17-segment AHA model, reflecting standardized myocardial perfusion abnormalities. Created in BioRender. Zhang, W. (2025) <https://BioRender.com/y27z326>. LSTM long short-term memory network, XGBoost extreme gradient boosting.

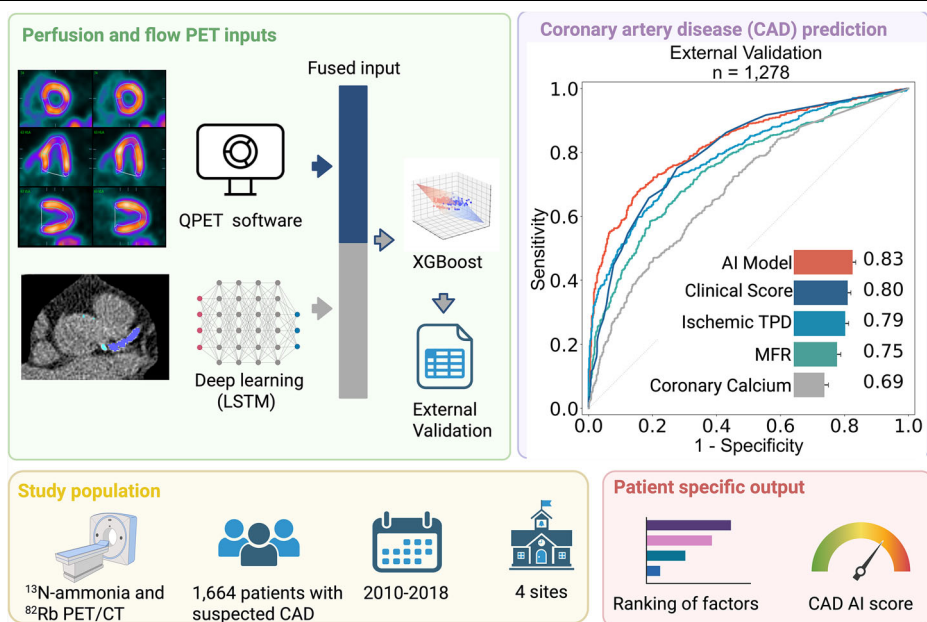
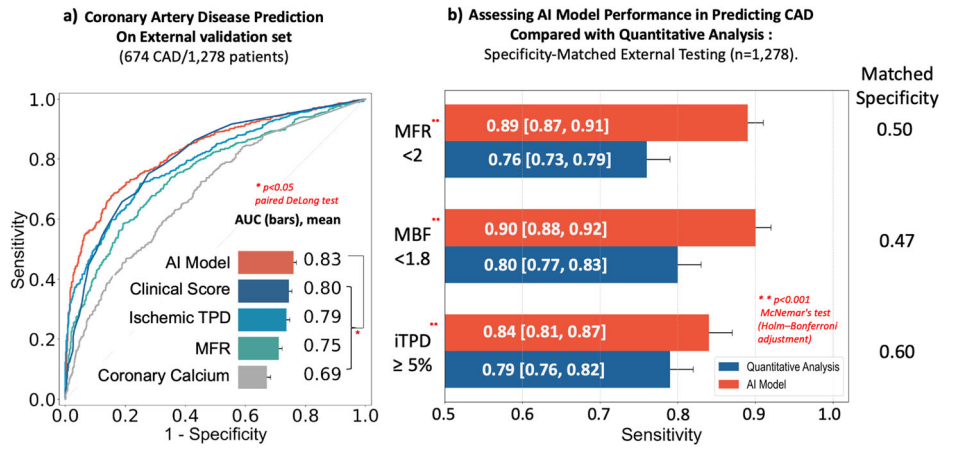


Fig. 2 | AI model performance. **a** Receiver-operating characteristics curve for coronary artery disease (CAD) and areas under the curve (AUC) values. AI Imaging model incorporates sex, coronary artery calcium score, rest and stress left ventricle ejection fraction (LVEF), rate pressure product (RPP), myocardial blood flow (MBF), and flow reserve (MFR), and perfusion measurements derived from PET imaging. **b** Sensitivity analysis comparing AI and quantitative methods for predicting CAD. Performance was assessed on the external validation set, with specificity matched between AI predictions and quantitative analysis. MFR and MBF were selected from the minimum of the 3 vessels. Notably, AI model shows significant improvement over MFR, MBF, and iTPD in terms of sensitivity. AI artificial intelligence, TID transient ischemic dilation, iTPD ischemic total perfusion deficit.



Few studies have attempted to improve the diagnostic accuracy of cardiac PET/CT through such an integrative approach^{3,6,15}. Poitrasson-Rivière et al. demonstrated that combining MFR and relative perfusion improves the detection of obstructive CAD⁴. Beyond flow and perfusion measures, assessments of the coronary atherosclerotic burden can provide additional information. Zampella et al. demonstrated that combining CAC, MBF, and MFR provides incremental information about the presence of CAD¹⁷. Similarly, Brodov et al. showed that integrating per-vessel ischemic TPD with CAC improves CAD detection¹³. While previous studies combining PET and CAC data relied on dedicated ECG-gated CT and manual expert analysis, in this study, we obtained CAC data automatically from PET CTAC. By leveraging CTAC for disease burden assessments, our approach can be employed for all PET MPI studies rather than being limited to exams that include a dedicated CAC scan. Our study is the first AI empowered multicenter PET analysis with rigorous external validation. While previous efforts relied on single-center data and hence employed prospective validation, we developed our model using data from one site and tested its performance on unseen data from three independent high-volume PET centers. This provides a critical evaluation of the external generalizability of our approach, which is further strengthened by the utilization of PET imaging measurements that vary less between institutions compared to actual PET images¹⁸.

In our study, XGBoost was compared against conventional logistic regression. Both models performed similarly in terms of discrimination and calibration, consistent with prior evidence that machine learning often does not outperform well-specified regression models when sample size and predictors are limited¹⁹. Nonetheless, XGBoost offers advantages that may become more relevant in larger cohorts or with richer feature set^{20–22}.

Traditional clinical studies combining multiple PET parameters required separate threshold values for key parameters and relied on conventional statistics^{3,12}. In contrast, our AI model integrates multi-dimensional data as continuous variables, delivering a probability-based assessment of obstructive CAD without relying on arbitrary thresholds. The results presented in this study suggest that a holistic AI-based approach improves overall diagnostic sensitivity and enables more accurate identification of high-risk patients who may be missed by individual quantitative metrics. Using SHAP values, the model highlights parameters influencing CAD predictions for individual patients (Fig. 3), enabling physicians to validate and understand case-specific diagnostic factors. This explainable AI model enhances clinical workflows by providing quantitative, automated, and objective insights, ultimately improving the efficiency of PET/CT MPI-based CAD prediction. By offering patient-specific diagnostic insights, it empowers physicians with comprehensive and actionable decision support at the point of care.

This study has a few limitations. While it was a retrospective study that only included four sites, it represents the largest PET/MPI and CAC study with invasive angiography data used as a gold standard. At this stage, the model is intended strictly for research use and is not approved for clinical deployment. Further external validation—ideally through prospective studies evaluating workflow integration and clinical impact—will be required before regulatory consideration and real-world translation. While most processing was automated, there was a quality control step of myocardial contours performed by an experienced technologist. Another limitation is the potential for referral bias, as only 10% of the overall PET MPI cohort underwent ICA within 180 days. These patients have higher rates of medical comorbidities and more abnormal perfusion findings compared to the overall PET MPI referral populations. However, these selection criteria are unavoidable in cases in which the presence or absence of obstructive CAD needs to be confirmed by invasive testing^{23,24}. Additionally, although invasive angiography serves as a traditional diagnostic endpoint, it is important to acknowledge the possibility of significant disease being present despite negative angiographic findings. For example, flow assessment by PET MPI can detect microvascular dysfunction and diffuse atherosclerosis without significant angiographic stenosis, an aspect that was not addressed in the current study. A limitation is that our reference standard was based solely on anatomical stenosis thresholds ($\geq 50\%$ left main, $\geq 70\%$ other vessels), as fractional flow reserve (FFR) and instantaneous wave-free ratio (iFR) measurements were not available in our cohort²⁵. This approach may not fully capture the physiological significance of coronary lesions or account for microvascular dysfunction. Prior studies have demonstrated substantial discordance between anatomical severity and functional ischemia, underscoring the importance of future work incorporating functional indices^{26,27}. Finally, race analysis was excluded from this study due to the limited availability of data on individuals from racial and ethnic minority groups, such as Black individuals.

This study introduces a novel AI model that integrates PET MPI parameters with deep learning-derived calcium scores, significantly enhancing CAD diagnosis. It outperforms experienced physicians, offers robust and interpretable assessments, and represents the first multicenter study with external validation for AI-driven cardiac PET MPI analysis.

Methods

Study population

In our multi-center study involving 17,348 patients undergoing cardiac PET/CT from the Registry of Fast Myocardial Perfusion Imaging with NExt generation PET (REFINE PET), we retrospectively enrolled 1664 patients from four sites who had suspected CAD, underwent ⁸²Rubidium or ¹³N-ammonia PET MPI, and ICA within 180 days from the PET/CT (Supplementary Fig. 5). CAD was defined as left main diameter stenosis

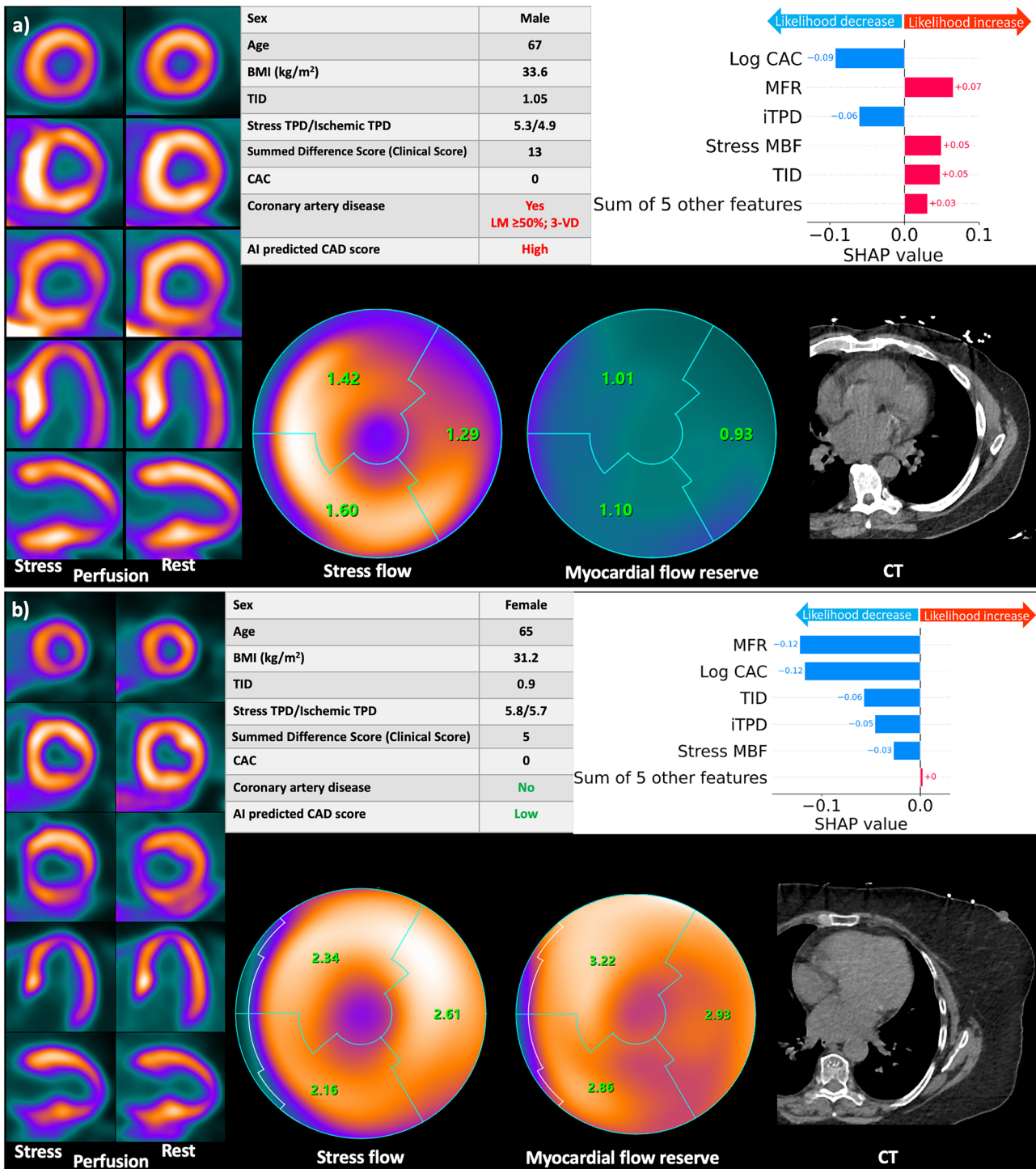


Fig. 3 | Examples of PET/CT imaging in predicted CAD patients. Examples of patients undergoing PET/CT myocardial perfusion imaging predicted to have CAD. Bar plots demonstrate the top 5 parameters with the highest influence on per-patient prediction of CAD. **a** Case where patient has abnormal MFR and stress MBF, borderline stress TPD, zero CAC, left main $\geq 50\%$ and $\geq 70\%$ in other three arteries (99% in LAD, 70% in distal LCX, and 70% in distal RCA); AI model estimated the likelihood of CAD to be high in agreement with the clinical score. **b** Case where patient has normal MFR and stress MBF, borderline ischemic and stress TPD, zero CAC, no actual CAD (20% stenosis in mid RCA; no stenosis elsewhere). The clinical

score indicated an abnormal scan. The AI model correctly estimated the likelihood of CAD to be low, aligning with the patient’s actual condition. While iTPD for the patient is borderline, the presence of other factors, such as normal blood flow, reduces the probability of the disease. BMI body mass index, CAC coronary calcium score, CAD coronary artery disease, CT computed tomography, LCX left circumflex artery, LAD left anterior descending, LM left main, LVEF left ventricular ejection fraction, MBF myocardial blood flow, MFR myocardial flow reserve, PET positron emission tomography, RPP rate pressure product, TID transient ischemic dilation, TPD total perfusion deficit, VD vessel disease.

$\geq 50\%$ or $\geq 70\%$ diameter stenosis in other epicardial arteries, based on clinical ICA readings and review of the images by an experienced physician. At each site, perfusion, quantitative flow, function, and CTAC data were interpreted by local readers, and reports were sent to the ordering physicians

to guide decisions on whether ICA was appropriate. Patients with prior myocardial infarction, percutaneous coronary intervention (PCI), and coronary artery bypass graft (CABG) were excluded from the analysis (Table 1). For patients who underwent several exams within the study

period, only the initial exam was considered. Data from a single site, comprising 386 patients, were used for model training and optimization, while data from three additional sites, totaling 1278 patients, were reserved for external testing. Clinical and demographic data were collected on the day of the MPI scan. All ICA referrals followed routine clinical practice, with decisions based on patients' symptom status, past medical history, and PET perfusion/flow findings from clinical reports, particularly the severity of flow reduction and perfusion abnormality²³.

Institutional review boards (IRB) approval was obtained at each site, and the study complies with the Declaration of Helsinki. Sites either obtained written informed consent or a waiver of consent for the use of the de-identified data.

PET protocol

For all patients, same-day rest and pharmacologic stress ⁸²Rubidium or ¹³N-ammonia PET MPI studies were conducted using a Biograph mCT 64 PET/CT scanner (Siemens Healthineers), a Discovery RX scanner (GE Healthcare), or a Discovery 710 scanner (GE Healthcare). A 6-min rest list-mode acquisition was started immediately before administering weight-based doses of ⁸²Rubidium or ¹³N-ammonia. Pharmacologic stress was induced using regadenoson ($n = 1489$), adenosine ($n = 131$), dobutamine ($n = 25$), or dipyridamole ($n = 19$). Concurrent with the beginning of the injection, a 6-min stress imaging acquisition was initiated. Prior to each PET acquisition for rest and stress, a low-dose helical CT scan was performed for attenuation correction, as detailed previously²⁸.

PET imaging and quantification

All PET imaging variables, including myocardial perfusion, blood flow, ejection fraction, and TID, were computed automatically in batch mode at the core laboratory with dedicated software (QPET, Cedars-Sinai Medical Center, Los Angeles, CA), ensuring standardized measurements between sites^{29–31}. Rest and stress relative perfusion were quantified using TPD. Ischemic TPD is defined as stress TPD—rest TPD; Normal myocardial perfusion was defined as stress TPD < 5%³². Rest and stress MBF were measured using a 1-tissue compartment kinetic model for ⁸²Rubidium PET and a 2-compartment model for ¹³N-ammonia PET^{29–31}. MBF and the spillover fraction from the blood to the myocardium were determined via numeric optimization. Stress and rest flow values, expressed in units of mL/g/min, were computed locally at each pixel on the polar map. Minimal segmental stress MBF (stress MBF) was used in the training of the AI model³³. Specifically, stress MBF and MFR were derived from the 17-segment AHA model using the lowest segmental value. The rate–pressure product (RPP) was derived by multiplying the heart rate (bpm) by the systolic blood pressure (mmHg), which was then applied to normalize the rest MBF within the angiographic group using the formula (rest MBF × average RPP)/RPP. Heart rate and blood pressure used for calculating the RPP were obtained at rest. The average RPP value in the angiographic group was 8500 bpm mmHg^{34,35}. MFR was computed in each segment as the ratio of stress MBF to rest MBF (rest flow adjusted by RPP) for the same segmental anatomy.

Clinical scoring

PET/CT scans were visually assessed during clinical reporting by experienced physicians at each site, with knowledge of all available data, including stress and rest perfusion imaging, gated functional data, myocardial flow reserve, all other quantitative information, CT images, and clinical information. The final visual assessment was performed by summed stress scores (SSS), summed rest scores (SRS), and summed difference score (SDS) using the 17-segment American Heart Association model. The SDS was used as the final clinical score due to its reliability and standardization in summarizing myocardial perfusion abnormalities.

AI coronary artery calcium scoring

We used our previously validated deep learning (DL) model for CAC segmentation and scoring^{2,36}. This model was trained and internally validated

on data from three centers, including a total of 9543 scans: 1827 electrocardiographically gated CAC scans and 7716 CTAC maps². Using the established deep learning segmentation method, we automatically derived CAC scores from CTAC maps.

Classification model and internal model testing

XGBoost models (version 1.7.3), a leading machine learning approach, were used for CAD diagnosis²⁰. Initially, 10-fold cross-validation was applied across the training dataset, where each fold allocated 90% of the data for model training and the remaining 10% for internal validation. To optimize model performance, hyper-parameter tuning was conducted within each fold through grid search, selecting the optimal hyper-parameter configuration from all tested combinations. In each 10-fold subset, an internal 10-fold cross-validation regimen further allocated 90% of the subset data for fitting and 10% for tuning, maximizing training data utilization and mitigating overfitting (Supplementary Fig. 6).

For the subsequent model evaluation phase, an external testing set was used. This additional testing set, originating from three distinct sites from that of the training data, was employed to rigorously assess the predictive performance on unseen data, providing an unbiased evaluation. The final model tested was created using the optimal hyperparameters obtained through grid search, retrained on the entire training set, and then tested on this external testing set.

The AI model outputs a continuous likelihood score between 0 and 1, with a higher value indicating a higher probability of being diagnosed as CAD. For binary classification, likelihoods were converted into binary outcomes using thresholds calibrated to the specificity of an established clinical metric (such as ischemic TPD). This threshold ensures the AI model identifies a similar proportion of patients as “normal,” allowing for clinically meaningful comparison between AI predictions and traditional assessments.

The diagnostic performance of the holistic AI model was evaluated using AUC and sensitivity analysis. Sensitivity was compared by adjusting prediction thresholds to match specificity across methods, then assessing the corresponding sensitivity. Comparisons focused on high-risk groups in the external testing set, defined by Duke category 6 (≥50% left main stenosis, ≥70% 3-vessel disease, or 2-vessel CAD involving proximal left anterior descending (LAD) and Duke category 5 (2-vessel severe stenosis excluding proximal LAD, one-vessel severe stenosis in proximal LAD, or ≥50% 3-vessel moderate stenosis)³⁷. Quantitative analysis includes minimum 3-vessel MFR < 2³², minimum 3-vessel stress MBF < 1.8³⁸, and ischemic TPD > 5%¹³.

Additionally, subgroup analyses were conducted to gain further insights into the performance of the AI models. Subgroups were categorized based on sex (male, female), age (<65, ≥65 years), and BMI (<30 kg/m², ≥30 kg/m²)³⁹. Due to limited data for racial and ethnic minorities, the race-based subgroup analysis was excluded.

Models

Six models were used for the CAD diagnosis: 1—CAC derived from CTAC maps, 2—ischemic TPD, 3—stress MBF, 4—MFR, 5—summed difference score (clinical score) obtained during clinical reading by an experienced physician, 6—AI, which employs PET/CT-derived data solely without incorporating additional clinical variables. The AI model incorporated the CAC along with: Stress MBF, MFR, ischemic TPD, stress TPD, stress and rest left ventricular ejection fraction (LVEF), RPP, TID, and sex, resulting in a total of 10 image-derived features to mimic clinical practice (Table 2). The model is designed to hold regardless of age, BMI, obesity, or other demographic factors.

Model explainability

The predictive power of model variables was assessed using XGBoost feature importance, measured by information gain to quantify accuracy improvement from each feature²⁰. Features were ranked by overall importance, with the most influential at the top. SHAP, a game-theory method, explained how features contributed to individual predictions⁴⁰.

In SHAP plots, dots represent the SHAP value of each patient for a feature, with positive values increasing predictions and negative values decreasing them. The x-axis distance from zero indicates the magnitude of influence.

Statistical analysis

We assessed the distribution of data using the Shapiro–Wilk test. Categorical variables were reported as n (%) and continuous variables with a normal distribution as mean \pm standard deviation (SD), while non-normal distribution was reported as median with IQR [IQ1–IQ3]. We used the Pearson's χ^2 test to assess the differences between categorical variables, the student's t -test for continuous variables with normal distribution, the Mann–Whitney–Wilcoxon test for non-parametric continuous variables. Predictions of CAD by ischemia, stress MBF, MFR, coronary calcium, and AI models were assessed by pairwise comparisons of the AUC with the DeLong test⁴¹. A two-tailed p -value of <0.05 was considered statistically significant. All statistical analyses were performed with Pandas (version 2.1.1), Numpy (version 1.24.3), Scipy (version 1.11.4), Lifelines (version 0.28.0) and Scikit-learn (version 1.3.0) in Python 3.11.5 (Python Software Foundation, Wilmington, DE, USA), as well as “nricens” package (version 1.6) in R version 4.3.2 (R Foundation for Statistical Computing, Vienna, Austria).

Data availability

To the extent allowed by data sharing agreements and IRB protocols, the deidentified data from this manuscript will be shared upon written request.

Code availability

The analysis code has been made available on GitHub: <https://github.com/qimagingAI/AI4CADdiagnosisPET-CT> under a Creative Commons BY-NC license. This model is intended for research use only and is not currently cleared for clinical deployment.

Received: 19 June 2025; Accepted: 1 January 2026;

Published online: 14 January 2026

References

- Di Carli, M. F., Dorbala, S. & Hachamovitch, R. Integrated cardiac PET-CT for the diagnosis and management of CAD. *J. Nucl. Cardiol.* **13**, 139–144 (2006).
- Pieszko, K. et al. Deep learning of coronary calcium scores from PET/CT attenuation maps accurately predicts adverse cardiovascular events. *JACC Cardiovasc. Imaging* **16**, 675–687 (2023).
- Murthy, V. L. et al. Improved cardiac risk assessment with noninvasive measures of coronary flow reserve. *Circulation* **124**, 2215–2224 (2011).
- Poitrasson-Rivière, A. et al. Integrated myocardial flow reserve (iMFR) assessment: optimized PET blood flow quantification for diagnosis of coronary artery disease. *Eur. J. Nucl. Med. Mol. Imaging* **51**, 136–146 (2023).
- McCubrey, R. O. et al. A highly predictive cardiac positron emission tomography (PET) risk score for 90-day and one-year major adverse cardiac events and revascularization. *J. Nucl. Cardiol.* **30**, 46–58 (2023).
- van Velzen, S. G. M. et al. Automated cardiovascular risk categorization through AI-driven coronary calcium quantification in cardiac PET acquired attenuation correction CT. *J. Nucl. Cardiol.* **30**, 955–969 (2023).
- Anderson, J. L. et al. Absent or mild coronary calcium predicts low-risk stress test results and outcomes in patients considered for Flecainide therapy. *J. Cardiovasc. Pharmacol. Ther.* **26**, 648–655 (2021).
- Maddahi, J. et al. Flurpiridaz F-18 PET myocardial perfusion imaging in patients with suspected coronary artery disease. *J. Am. Coll. Cardiol.* **82**, 1598–1610 (2023).
- Gupta, A. et al. Integrated noninvasive physiological assessment of coronary circulatory function and impact on cardiovascular mortality in patients with stable coronary artery disease. *Circulation* **136**, 2325–2336 (2017).
- Mannarino, T. et al. Combined evaluation of CAC score and myocardial perfusion imaging in patients at risk of cardiovascular disease: where are we and what do the data say. *J. Nucl. Cardiol.* **30**, 2349–2360 (2023).
- Reeves, R. A., Halpern, E. J. & Rao, V. M. Cardiac imaging trends from 2010 to 2019 in the medicare population. *Radiol. Cardiothorac. Imaging* **3**, e210156 (2021).
- Ziadi, M. C. et al. Impaired myocardial flow reserve on rubidium-82 positron emission tomography imaging predicts adverse outcomes in patients assessed for myocardial ischemia. *J. Am. Coll. Cardiol.* **58**, 740–748 (2011).
- Brodov, Y. et al. Combined quantitative assessment of myocardial perfusion and coronary artery calcium score by hybrid 82Rb PET/CT improves detection of coronary artery disease. *J. Nucl. Med.* **56**, 1345–1350 (2015).
- Gould, K. L. et al. Mortality prediction by quantitative PET perfusion expressed as coronary flow capacity with and without revascularization. *JACC Cardiovasc. Imaging* **14**, 1020–1034 (2021).
- Singh, A. et al. Deep learning for explainable estimation of mortality risk from myocardial positron emission tomography images. *Circ. Cardiovasc. Imaging* **15**, e014526 (2022).
- Valenta, I. & Schindler, T. H. PET-determined myocardial perfusion and flow in coronary artery disease characterization. *J. Med. Imaging Radiat. Sci.* **55**, S44–S50 (2024).
- Zampella, E. et al. Combined evaluation of regional coronary artery calcium and myocardial perfusion by (82)Rb PET/CT in predicting lesion-related outcome. *Eur. J. Nucl. Med. Mol. Imaging* **47**, 1698–1704 (2020).
- Akamatsu, G. et al. A review of harmonization strategies for quantitative PET. *Ann. Nucl. Med.* **37**, 71–88 (2023).
- Christodoulou, E. et al. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J. Clin. Epidemiol.* **110**, 12–22 (2019).
- Chen, T. & Guestrin, C. XGBoost: a scalable tree boosting system. In: Balaji K et al. (eds) *Proc. of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (Association for Computing Machinery, 2016).
- van Calster, B. et al. Calibration: the Achilles heel of predictive analytics. *BMC Med* **17**, 230 (2019).
- Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* **29**, 1189–1232 (2001).
- Patel, K. K. et al. Extent of myocardial ischemia on positron emission tomography and survival benefit with early revascularization. *J. Am. Coll. Cardiol.* **74**, 1645–1654 (2019).
- Thomas, M. et al. Reasons for discordance between positron emission tomography (PET) myocardial perfusion imaging (MPI) results and subsequent management. *J. Nucl. Cardiol.* **29**, 1109–1116 (2022).
- Yonetsu, T. et al. Significance of microvascular function in visual–functional mismatch between invasive coronary angiography and fractional flow reserve. *J. Am. Heart Assoc.* **6**, e005916 (2017).
- De Bruyne, B. et al. Fractional flow reserve-guided PCI versus medical therapy in stable coronary disease. *N. Engl. J. Med.* **367**, 991–1001 (2012).
- Tonino Pim, A. L. et al. Fractional flow reserve versus angiography for guiding percutaneous coronary intervention. *N. Engl. J. Med.* **360**, 213–224 (2009).
- Bacharach, S. L. PET/CT attenuation correction: breathing lessons. *J. Nucl. Med.* **48**, 677–679 (2007).
- Dekemp, R. A. et al. Multisoftware reproducibility study of stress and rest myocardial blood flow assessed with 3D dynamic PET/CT and a 1-tissue-compartment model of 82Rb kinetics. *J. Nucl. Med.* **54**, 571–577 (2013).

30. Nakazato, R. et al. Automated quantitative Rb-82 3D PET/CT myocardial perfusion imaging: normal limits and correlation with invasive coronary angiography. *J. Nucl. Cardiol.* **19**, 265–276 (2012).
31. Slomka, P. J. et al. Comparison of clinical tools for measurements of regional stress and rest myocardial blood flow assessed with ¹³N-ammonia PET/CT. *J. Nucl. Med.* **53**, 171–181 (2012).
32. Otaki, Y. et al. 5-Year prognostic value of quantitative versus visual MPI in subtle perfusion defects: results from REFINE SPECT. *JACC Cardiovasc. Imaging* **13**, 774–785 (2020).
33. Otaki, Y. et al. Improved myocardial blood flow estimation with residual activity correction and motion correction in ¹⁸F-flurpiridaz PET myocardial perfusion imaging. *Eur. J. Nucl. Med. Mol. Imaging* **49**, 1881–1893 (2022).
34. Kuronuma, K. et al. Automated motion correction for myocardial blood flow measurements and diagnostic performance of (⁸²Rb) PET myocardial perfusion imaging. *J. Nucl. Med.* **65**, 139–146 (2024).
35. Murthy, V. L. et al. Clinical quantification of myocardial blood flow using PET: joint position paper of the SNMMI cardiovascular council and the ASNC. *J. Nucl. Med.* **59**, 273–293 (2018).
36. Williams, M. C. et al. Automated vessel-specific coronary artery calcification quantification with deep learning in a large multi-centre registry. *Eur. Heart J. Cardiovasc. Imaging* **25**, 976–985 (2024).
37. Reynolds, H. R. et al. Outcomes in the ISCHEMIA trial based on coronary artery disease and ischemia severity. *Circulation* **144**, 1024–1038 (2021).
38. Hajjiri, M. M. et al. Comparison of positron emission tomography measurement of adenosine-stimulated absolute myocardial blood flow versus relative myocardial tracer content for physiological assessment of coronary artery stenosis severity and location. *JACC Cardiovasc. Imaging* **2**, 751–758 (2009).
39. Purnell, J. Q. *Definitions, Classification, and Epidemiology of Obesity* (MDText.com, Inc., 2000).
40. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. In: Ulrike von L et al. (eds) *Proc. of the 31st International Conference on Neural Information Processing Systems* 4768–4777 (Curran Associates Inc., 2017).
41. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44**, 837–845 (1988).

Acknowledgements

This research was supported in part by grant R35HL161195 from the National Heart, Lung, and Blood Institute of the National Institutes of Health and grant R01EB034586 from the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author contributions

W.Z. conducted data processing/experiments/analysis, and co-wrote the manuscript. J.K. co-wrote the manuscript and contributed to materials and clinical expertise. P.J.S. designed the study, provided the overall guidance and study funding, co-wrote the manuscript, and contributed materials, clinical expertise, and technical expertise. A.S., R.J.M., S.M., G.R., J.Y., D.H., D.D., G.D., G.K., M.L., P.B.K., J.X.L., J.Z., V.B., J.H., S.C., L.B., A.J.E., S.K., S.M., V.T.L., W.A., S.W., P.C., D.S.B., and M.F.D.C. contributed to materials, clinical expertise, and technical expertise. All authors critically revised the

manuscript and contributed to its formation. All authors had full access to all the data in the study, accepted the final responsibility to submit for publication, and took responsibility for the contents of the manuscript.

Competing interests

Robert J.H. Miller received consulting fees and research support from Pfizer. Daniel S. Berman, Piotr J. Slomka, and Paul B. Kavanagh participated in software royalties for QPS software at Cedars-Sinai Medical Center. Piotr S. Slomka has received consulting fees from Synektik. Damini Dey, Piotr J. Slomka and Daniel S. Berman have equity interest in APQ Health. Daniel S. Berman has served as a consultant for GE Healthcare. Andrew J. Einstein has received speaker fees from Ionetix, consulting fees from Artrya and W. L. Gore & Associates, and authorship fees from Wolters Kluwer Healthcare. Andrew J. Einstein has also served on scientific advisory boards for Canon Medical Systems and Synektik S.A. and received grants to Columbia University from Alexion, Attralus, BridgeBio, Canon Medical Systems, Eidos Therapeutics, Intellia Therapeutics, International Atomic Energy Agency, Ionis Pharmaceuticals, National Institutes of Health, Neovasc, Pfizer, Roche Medical Systems, Shockwave Medical, and W. L. Gore & Associates. Marcelo F. Di Carli reports on institutional research grants from Gilead Sciences, Xylocor, Sun Pharma, Intellia Therapeutics, Alnylam Pharmaceuticals, and Amgen. He also receives consulting fees from MedTrace, Valo Health, and IBA. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41746-026-02338-6>.

Correspondence and requests for materials should be addressed to Piotr J. Slomka.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

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

© The Author(s) 2026