Published in final edited form as:
Eur J Nucl Med Mol Imaging. 2023 January ; 50(2): 387–397. doi:10.1007/s00259-022-05972-w.

Mitigating bias in deep learning for diagnosis of coronary artery disease from myocardial perfusion SPECT images

Robert JH Miller, MD[1,2,*], Ananya Singh, MS[1,*], Yuka Otaki, MD, PhD[1], Balaji K. Tamarappoo, MD[1], Paul Kavanagh, MS[1], Tejas Parekh, BSc[1], Lien-Hsin Hu, MD[1,3], Heidi Gransar, MS[1], Tali Sharir, MD[4], Andrew J. Einstein, MD, PhD[5], Mathews B. Fish, MD[6], Terrence D. Ruddy, MD[7], Philipp A. Kaufmann, MD[8], Albert J. Sinusas, MD[9], Edward J. Miller, MD PhD[9], Timothy M. Bateman, MD[10], Sharmila Dorbala, MD, MPH[11], Marcelo Di Carli, MD[11], Joanna X. Liang, BA[1], Damini Dey, PhD[1], Daniel S. Berman, MD[1], Piotr J. Slomka, PhD[1]

[1]Departments of Medicine (Division of Artificial Intelligence), Imaging and Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA, USA

[2]Department of Cardiac Sciences, University of Calgary, Calgary AB, Canada

[3]Department of Nuclear Medicine, Taipei Veterans General Hospital, Taipei, Taiwan

[4]Department of Nuclear Cardiology, Assuta Medical Centers, Tel Aviv, Israel and Ben Gurion University of the Negev, Beer Sheba, Israel

[5]Division of Cardiology, Department of Medicine, and Department of Radiology, Columbia University Irving Medical Center and New York-Presbyterian Hospital, New York, NY, USA

[6]Oregon Heart and Vascular Institute, Sacred Heart Medical Center, Springfield, OR, USA

[7]Division of Cardiology, University of Ottawa Heart Institute, Ottawa, ON, Canada

[8]Department of Nuclear Medicine, Cardiac Imaging, University Hospital Zurich, Zurich, Switzerland

[9]Section of Cardiovascular Medicine, Department of Internal Medicine, Yale University School of Medicine, New Haven, CT, USA

[10]Cardiovascular Imaging Technologies LLC, Kansas City, MO, USA

[11]Department of Radiology, Division of Nuclear Medicine and Molecular Imaging, Brigham and Women's Hospital, Boston, MA, USA

Address for Correspondence: Piotr Slomka, PhD, Cedars-Sinai Medical Center, 8700 Beverly Boulevard, Ste. Metro 203, Los Angeles, California 90048, Phone: 310-423-4348 Fax: 310-423-0173, Piotr.Slomka@cshs.org.

Author Contributions

All authors contributed to the study conception and design. Material preparation and data collection were performed by and analysis were performed by Piotr J. Slomka, Tali Sharir, Andrew J. Einstein, Mathews B. Fish Terrence D. Ruddy, Philipp A. Kaufmann, Albert J. Sinusas, Edward J. Miller, Timothy M. Bateman, Sharmila Dorbala, and Marcelo Di Carli. The first draft of the manuscript was written by Robert J.H. Miller, Ananya Singh, and Yuka Otaki. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

[*]Authors contributed equally

Ethics Approval

The study protocol complied with the Declaration of Helsinki and was approved by the institutional review boards at each participating institution and the overall study was approved by the institutional review board at Cedars-Sinai Medical Center.

Abstract

Purpose: Artificial intelligence (AI) has high diagnostic accuracy for coronary artery disease (CAD) from myocardial perfusion imaging (MPI). However, when trained using high-risk populations (such as patients with correlating invasive testing), the disease probability can be overestimated due to selection bias. We evaluated different strategies for training AI models to improve the calibration (accurate estimate of disease probability), using external testing.

Methods: Deep learning was trained using 828 patients from 3 sites, with MPI and invasive angiography within 6-months. Perfusion was assessed using upright (U-TPD) and supine total perfusion deficit (S-TPD). AI training without data augmentation (Model 1) was compared to training with augmentation (increased sampling) of patients without obstructive CAD (Model 2), and patients without CAD and TPD <2% (Model 3). All models were tested in an external population of patients with invasive angiography within 6 months (n=332) or low likelihood of CAD (n=179).

Results: Model 3 achieved the best calibration (Brier score 0.104 vs 0.121, p<0.01). Improvement in calibration was particularly evident in women (Brier score 0.084 vs 0.124, p<0.01). In external testing (n=511), the area under the receiver operating characteristic curve (AUC) was higher for Model 3(0.930), compared to U-TPD(AUC 0.897) and S-TPD(AUC 0.900, p<0.01 for both).

Conclusion: Training AI models with augmentation of low-risk patients can improve calibration of AI models developed to identify patients with CAD, allowing more accurate assignment of disease probability. This is particularly important in lower-risk populations and in women, where overestimation of disease probability could significantly influence down-stream patient management.

Keywords

deep learning; model training; calibration; diagnostic accuracy; sex-specific analysis

INTRODUCTION

Coronary artery disease (CAD) is a major cause of death among both men and women [1, 2]. Single-photon emission computed tomography (SPECT) myocardial perfusion imaging (MPI) is one of the most common imaging modalities used for assessment of ischemia [3]. Previously, we demonstrated improved prediction of obstructive CAD with an artificial intelligence (AI) deep learning from raw, extent, and quantitative polar maps from SPECT MPI [4, 5]. We recently enhanced the model to utilize only raw perfusion polar maps (removing reliance on specific software), but with additional clinical information (age, sex, and cardiac volumes), which can be obtained automatically from image files[6]. This algorithm also incorporates gradient-weighted Class Activation Mapping (Grad-CAM) [7] to produce a coarse localization map highlighting the important regions in the image for prediction, providing explainability for this AI approach[6].

However, a fundamental issue for developing any AI model is ensuring that the training population reflects the population where the model will ultimately be applied. This is potentially problematic for models trained to predict the presence of CAD. Patients who have undergone both MPI and invasive coronary angiography (ICA) are typically used for training and testing populations but are a highly selected population with a high prevalence of CAD and more abnormal MPI findings compared to populations where AI will ultimately be applied. This could lead to an inaccurate estimate of probability particularly in lower-risk subgroups [8]. Calibration, a measure of how closely predicted probability reflects actual probability, is critical when diagnosing obstructive CAD since this probability directly impacts physician decision-making regarding referral for invasive testing[9]. Although poor calibration is potentially an issue for AI models, the impact of different methods for training data augmentation (expanding training data samples from the existing dataset) on model accuracy and calibration has not been extensively studied previously.

In this study, we evaluated the influence of different methods for enhancing training with data augmentation on the diagnostic accuracy and calibration of a SPECT MPI AI model for predicting presence of obstructive CAD. The final evaluation was performed in an external population, with specific attention to performance in both women and men.

MATERIALS AND METHODS

Study Population

The studied dataset was collected under NIH-sponsored REgistry of Fast Myocardial Perfusion Imaging with NExt generation SPECT (REFINE SPECT) [10]. The diagnostic registry contains MPI studies of consecutive patients without known CAD, who underwent clinically indicated ICA within 180 days of MPI [5]. In this study, we included patients who underwent imaging with a DSPECT camera system (Spectrum Dynamics, Caesarea, Israel) with both supine and upright image acquisitions. We have additionally included a random subset of patients from the prognostic registry who underwent clinically indicated SPECT MPI with a DSPECT camera system, without ICA but with a low-likelihood (LLK) of CAD, described previously[11]. This population was divided into a training population (n=828 from three sites) and an external testing population (n=511 from a separate site). The primary endpoint for the study was prediction of obstructive CAD. To the extent allowed by data sharing agreements and institutional review board protocols, the data from this manuscript will be shared upon written request. The study protocol complied with the Declaration of Helsinki and was approved by the institutional review boards at each participating institution, and the overall study was approved by the institutional review board at Cedars-Sinai Medical Center.

Training population

The training population included MPI images from 828 patients from three sites who underwent upright and supine SPECT MPI between 2008 and 2015. We additionally included patients with LLK of CAD (n = 186) from the same sites, which was assumed to be equivalent to no obstructive CAD, for augmentation. The criteria for LLK included: pre-test probability of CAD <0.10 [12], no history of CAD [13], no diabetes or peripheral vascular disease, left ventricular ejection fraction ≥50% and MPI interpreted visually as normal.

External Testing population

Testing population included 511 patients from a separate site which either underwent ICA within 180 days (n=332) or had a LLK of CAD (n=179). The external site was selected randomly while ensuring a population of LLK cases could be identified.

Stress and Acquisition Protocols

SPECT MPI was performed using Tc-99m sestamibi on D-SPECT (Spectrum-Dynamics, Israel) scanners [14]. Patients underwent either symptom-limited Bruce protocol treadmill exercise testing stress or pharmacologic stress, with radiotracer injection at peak exercise or during maximal hyperemia, respectively. Upright and supine stress imaging began 15–60 min after stress, with acquisitions occurring over 4–6 minutes. No attenuation, scatter, or motion correction was applied.

Coronary Angiography

ICA was performed according to standard clinical protocols. All coronary angiograms were visually interpreted by an on-site cardiologist. Obstructive CAD was defined as luminal diameter narrowing of 50% or greater in the left main artery, or 70% or greater in the left anterior descending artery (LAD), left circumflex artery (LCx), or right coronary artery (RCA).

Visual Perfusion Assessment

Summed stress score (SSS), using a 17-segment model [15, 16], was assessed during clinical reporting by experienced board-certified nuclear cardiologists at each site with knowledge of all available data, including quantitative perfusion (supine, and upright), gated functional data, and all clinical information according to routine local protocols.

Automated Image Quantification

Myocardial contours were generated automatically with quantitative perfusion SPECT (QPS)/Quantitative Gated SPECT (QGS) software (Cedars-Sinai Medical Center, Los Angeles, CA, USA) and when necessary, contours were adjusted by an experienced core laboratory technologist [10]. Total perfusion deficit (TPD) was automatically generated as previously described [17]. TPD was derived from stress acquisition in the upright position (U-TPD), and supine position (S-TPD). TPD was generated by automated processing with sex-specific normal limits. Ejection fraction, end-systolic and end-diastolic volume at stress and rest were derived using QGS software. For comparison of false positive rates, S-TPD ≥ 3% was used as the threshold for abnormal [18].

Deep Learning and Grad-CAM Implementation

The AI model was a convolutional neural network which incorporated supine and upright raw perfusion maps as the inputs in polar coordinates and is therefore not reliant on any specific software package. Age and sex were extracted automatically from Digital Imaging and Communications in Medicine (DICOM) image headers. Cardiac volumes (stress end-diastolic and end-systolic volumes) were quantified automatically from rest and stress gated images (Figure 1). The output of the AI model is a 1×4 probability vector of per vessel (for LAD, LCx and RCA) and per-patient obstructive CAD prediction. Grad-CAM[7] was incorporated to highlight regions of polar maps contributing to the prediction and presented as an attention map, which highlights the polar map areas contributing to predictions, as well as a CAD probability map which shows segments contributing to predictions and per-vessel probability of CAD. A case example is shown in Supplemental Figure 1. The model was implemented using Python 3.7.3 and Tensorflow 1.14. The training was performed using Titan RTX graphics card (Nvidia, Santa Clara CA). Further details about data pre-processing and model architecture are provided in Supplemental Material.

Calibration of Deep Learning Analysis with Augmentation

To calibrate the AI model for real world data, we tested 3 methods of training. Model 1, our previous training method[6], was trained without any data augmentation but with a weighted loss function to account for class imbalance. Model 2 was trained with augmenting data in patients without obstructive CAD. Model 3 was trained with augmenting "simulated low likelihood" data defined as patients with angiographically normal coronary arteries and S-TPD < 2%. Polar maps were rotated +/− 10 degrees for augmentation. Histogram and calibration plots as well as Brier scores for all three models were compared. Differences in calibration were assessed with t-tests of the squared error [19].

Model Training

To train the model, data from the 3 training sites were split in 5 folds randomly, with 2 folds (20% each) held out for validation and testing in each fold and the rest for training. An equal proportion of patients with obstructive CAD were maintained in each of the folds. This process was repeated 5 times, with each nonoverlapping fold used as test set, thus reducing variance in estimates caused by arbitrarily splitting the dataset once [20].

Our baseline model (Model 1) utilized weighted loss function to account for class imbalance (62.9% of patients with obstructive CAD) but did not use training data augmentation. We compared this method to a model trained with additional cases without significant CAD (Model 2) and with additional cases from simulated LLK patients (angiographically normal coronaries and S-TPD<2%, [Model 3]), to ensure balanced training folds (50% of patients with obstructive CAD).

External Testing

We used a separate site for external testing, which was not used in any way during model training. The model with the lowest validation loss (higher validation loss is suggestive of model over-fitting) during internal cross-validation was selected for external testing.

Statistical Analysis

Categorical variables are presented as number (frequency) and continuous variables as mean ± SD or median and interquartile ranges as appropriate. Variables were compared using a χ^2 statistic for categorical variables and a Wilcoxon rank-sum test for continuous variables. The diagnostic performance of SSS, stress TPD, and the AI model was evaluated using the receiver operating characteristic (ROC) analysis and pairwise comparisons of the area under the ROC curve (AUC) [21]. Thresholds were established in the training population to meet 90% sensitivity, then the same values were applied in the external testing population. A two-tailed P-value <0.05 was considered statistically significant. STATA version 14 was used for all analyses (Stata Corp, College Station, TX).

RESULTS

Population Characteristics

Patient characteristics from the training and external testing populations are shown in Table 1. Patients in the training population were older (mean age 64.0 vs 60.8, p<0.001) and the proportion of patients with CAD risk factors other than smoking was higher in the training population.

Angiographic Characteristics

In total, 521 (62.9%) patients in the training population and 197 (59.3%) patients from the testing population who underwent ICA had obstructive CAD. Angiographic characteristics of the training and testing population are shown in Table 2. There were no significant differences in the distribution of angiographic CAD between the two groups. Characteristics of patients undergoing ICA compared to LLK patients are shown in Supplemental Table 1.

Influence of Training Augmentation on Calibration

Model 3 had the best calibration (Brier score 0.104), which was significantly better compared to Model 1 (Brier score 0.121, p=0.003). Model 2 (Brier Score 0.108) had similar calibration to Model 3 (p=0.294). Calibration graphs for the overall population are shown in Figure 2. Model 1 tended to overestimate the probability of obstructive CAD in lower-risk patients (lowest 6 deciles). Correlation between predicted and actual probabilities were better in low-risk patients for Models 2 and 3. Model 3 demonstrated better calibration in women (Brier score 0.084) compared to men (Brier score 0.129) (Figure 3), while Model 1 showed similar calibration in men (Brier score 0.121) and women (Brier score 0.124). There were no differences in calibration when only considering patients who underwent ICA (Supplemental Figure 2). The calibration for model 3 (measured using Brier score) was not significantly worse in patients with elevated BMI compared to patients without (0.116 vs. 0.094, p=0.133), results in Supplemental Table 2. Additionally, the prediction performance for obstructive CAD was not significantly different for model 3 in patients with or without elevated BMI (AUC 0.913 vs 0.947, p=0.148).

False positive and false negative rates for each of the models are shown in Figure 4. Thresholds were derived in the training population to achieve 90% sensitivity. The false positive rate was significantly lower for Model 3 (21.2%) compared to Model 1 (31.5%, p=0.012). The proportion of abnormal tests in LLK patients was lower for both Model 2 (2.8%) and Model 3 (3.9%) compared to Model 1 (10.6%, p<0.05 for both). Model 3 had sensitivity of 88% and specificity 84% in the external population.

Diagnostic accuracy in the entire external population, including patients who underwent ICA and LLK patients, is shown in Figure 5. Model 3 had the highest diagnostic accuracy (AUC 0.930, 95% CI 0.908 – 0.952) with no significant differences between models. Only Model

3 had higher prediction performance compared to S-TPD (p=0.010 in overall population, p=0.011 in ICA population). Diagnostic accuracy for CAD in women and men is shown in Supplemental Figure 3. Predictive performance of the 3 models across the 5 validation folds is shown in Supplemental Table 3.

Prediction Performance Compared to Standard Quantification

Diagnostic accuracy in the external testing population, including patients who underwent ICA and LLK patients, is shown in Supplemental Figure 4. The AUC for the AI model (0.930), was better compared to U-TPD (AUC 0.897) and S-TPD (AUC 0.900, p<0.01 for both). Overall diagnostic accuracy was lower in the subset of patients who underwent ICA. However, the AUC for the AI model (0.877) was still higher compared to S-TPD (AUC 0.830, p=0.011) and U-TPD (AUC 0.829, p=0.002).

DISCUSSION

Using a large multi-center international registry, we demonstrate that training data augmentation has a significant impact on AI model calibration (reflecting model bias) and may also impact overall diagnostic accuracy. Models trained with additional cases from patients without obstructive CAD more closely predicted the probability of obstructive CAD in women and low-risk patients, which could potentially lead to less down-stream invasive testing. All AI models had similar diagnostic accuracy and outperformed quantitative analysis of perfusion when tested in an external population. Our results suggest that training data augmentation is critical to ensuring that AI predictions more closely reflect the population in which they will be applied.

AI models which are trained to predict the presence of obstructive CAD are limited by the use of training populations with inherent selection bias. Since the extent of CAD needs to be known, training populations must have undergone either ICA or coronary computed tomographic angiography. These populations have a higher prevalence of obstructive CAD and more abnormal perfusion compared to unselected populations referred for MPI. As a result, AI models trained on such populations tend to overestimate the probability of CAD in lower-risk patients. Two methods to overcome this issue are to use weighted loss functions or to augment training populations. With weighted loss functions, prediction errors in the under-represented population are assigned greater weight[22] (Model 1). Alternatively, training can be augmented with additional cases from the under-represented populations[23], as we did for Model 2 and Model 3. This allows the model to learn from additional cases without obstructive CAD, or in the case of Model 3, additional cases without CAD with near normal perfusion. These cases may better reflect the actual population of patients referred for MPI given the declining frequency of abnormal perfusion findings [24].

Our primary aim when assessing training data augmentation methods was to improve model calibration for prediction of obstructive CAD, when applied in a population which better reflected all patients referred for MPI. Model 1 tended to overestimate the probability of CAD in low-risk patients despite using a weighted loss function. Models 2 and 3 more closely predicted the probability of CAD in these low-risk patients, with a resultant lower false positive rate. Importantly, improvement in calibration was most evident in women which was an expected finding given the lower prevalence of obstructive CAD in women compared to men[25]. This result highlights the importance of assessing model discrimination and calibration by sex. When tested only in patients who underwent ICA, a population with similar selection bias as the training population, model calibration was similar regardless of the training method. These findings suggest that training augmentation is an effective way to improve model calibration without sacrificing diagnostic accuracy or calibration for high-risk patients.

In the overall population, the AI diagnostic accuracy was always higher compared to visual assessment or stress TPD on two positions for all models. Additionally, Model 3 had higher accuracy compared to S-TPD in the subset of patients with ICA. Model 3 also had higher prediction performance for identifying patients with obstructive CAD compared to our previous model incorporating upright and supine images (AUC 0.81)[26]. The present model also demonstrates higher AUC compared to a model integrating attenuation corrected and non-attenuation corrected imaging[27]. Importantly, the present model incorporates methods for explainable predictions, which we recently demonstrated could be used to improve the accuracy of physician interpretations [28]. Explainability is critical since, at least for the foreseeable future, AI models will be operating under physician supervision. In our study, there was also a trend towards higher diagnostic accuracy of Model 3 compared to S-TPD in both men and women. Only a few previous studies have evaluated the difference in the diagnostic accuracy of visual assessment for SPECT-MPI between men and women. A meta-analysis of 1,148 men and 1,142 women from 26 Anger SPECT studies showed that there was no significant difference in the sensitivity or specificity of visual perfusion assessment between men and women [26]. The mean sensitivity and specificity were 84.2% and 78.7% in women and 89.1% and 71.2% in men for the diagnosis of CAD using a cut-off of ≥50% stenosis [26]. In a small study of 61 women and 248 men, the accuracy of visual assessment of SPECT-MPI for detecting obstructive CAD with the CZT-camera in women was shown to be comparable to men [29]. To our knowledge, our study is the first to compare the diagnostic performance of AI to visual assessment and state-of-the-art automated quantification detected with a CZT-camera separately in men and women.

The study has some limitations. This was a retrospective analysis of patients who had undergone SPECT and ICA within 6 months. However, the actual interval between studies was relatively short (24±38 days). The degree of stenosis was visually assessed on invasive angiography and quantitative angiography was not performed in most sites. A stenosis ≥70% was used as a surrogate marker of hemodynamically significant stenosis. This study was performed using images from one camera system and the generalizability of our findings to other SPECT scanners remains to be evaluated. However, the augmentation of training populations was based on quantification of standard supine imaging and could potentially be applied more broadly. In this study, we assessed the impact of training augmentation on prediction performance and calibration in women and men. Future studies could evaluate whether these measures differ significantly by other patient features such as ethnicity or medical history. Additionally, while we demonstrated superior diagnostic accuracy for the AI model compared to expert visual interpretation and quantitative analysis of perfusion, prospective studies are needed to determine if this leads to a change in patient management and improved clinical outcomes. Lastly, we only evaluated three methods for training augmentation and there are likely an endless number of variations. However, our study does demonstrate the potential for training augmentation to improve calibration.

CONCLUSION

Augmenting training populations can lead to improved AI model calibration, and consequently more accurate assignment of the probability of the disease without impacting overall diagnostic accuracy. In our study this led to lower false positive rates, suggesting that it could significantly influence down-stream patient management.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding

This research was supported in part by grants R01HL089765 and R35HL161195 grant R01HL089765 from the National Heart, Lung, and Blood Institute/ National Institutes of Health (NHLBI/NIH) (PI: Piotr Slomka). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Competing Interests

Drs. Berman and Slomka and Mr. Kavanagh participate in software royalties for QPS software at Cedars-Sinai Medical Center. Dr. Slomka has received research grant support from Siemens Medical Systems. Drs. Berman is serving as a consultant for GE Healthcare (radiopharmaceuticals). Dorbala, Einstein, and Edward Miller have served as consultants for GE Healthcare. Dr. Einstein has served as a consultant to W. L. Gore & Associates. Dr. Dorbala has served as a consultant to Bracco Diagnostics; her institution has received grant support from Astellas. Dr. Di Carli has received research grant support from Spectrum Dynamics and consulting honoraria from Sanofi and GE Healthcare. Dr. Ruddy has received research grant support from GE Healthcare and Advanced Accelerator Applications. Dr. Einstein's institution has received research support from GE Healthcare, Philips Healthcare, Toshiba America Medical Systems, Roche Medical Systems, and W. L. Gore & Associates. The remaining authors have nothing to disclose.

Abbreviations

AI	artificial intelligence
AUC	area under the ROC curve
CAD	coronary artery disease
ICA	invasive coronary angiogram
LAD	left anterior descending coronary artery
LCx	left circumflex coronary artery
MPI	myocardial perfusion imaging
QPS	quantitative perfusion SPECT
RCA	right coronary artery
ROC	receiver operating characteristic

SPECT	single-photon emission computed tomography
TPD	total perfusion deficit

REFERENCES

1. Global Burden of Disease Collaborators. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017;390:1151–210. doi:10.1016/S0140-6736(17)32152-9. [PubMed: 28919116]
2. Townsend N, Wilson L, Bhatnagar P, Wickramasinghe K, Rayner M, Nichols M. Cardiovascular disease in Europe: epidemiological update 2016. Eur Heart J. 2016;37:3232–45. doi:10.1093/eurheartj/ehw334. [PubMed: 27523477]
3. Knuuti J, Wijns W, Saraste A, Capodanno D, Barbato E, Funck-Brentano C, et al. 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes. Eur Heart J. 2019;41:407–77. doi:10.1093/eurheartj/ehz425.
4. Betancur J, Hu LH, Commandeur F, Sharir T, Einstein AJ, Fish MB, et al. Deep Learning Analysis of Upright-Supine High-Efficiency SPECT Myocardial Perfusion Imaging for Prediction of Obstructive Coronary Artery Disease: A Multicenter Study. J Nucl Med. 2019;60:664–70. doi:10.2967/jnumed.118.213538. [PubMed: 30262516]
5. Betancur J, Commandeur F, Motlagh M, Sharir T, Einstein AJ, Bokhari S, et al. Deep Learning for Prediction of Obstructive Disease From Fast Myocardial Perfusion SPECT: A Multicenter Study. JACC Cardiovasc Imaging. 2018;11:1654–63. doi:10.1016/j.jcmg.2018.01.020. [PubMed: 29550305]
6. Otaki Y, Singh A, Kavanagh P, Miller RJH, Parekh T, Tamarappoo BK, et al. Clinical Deployment of Explainable Artificial Intelligence of SPECT for Diagnosis of Coronary Artery Disease. JACC Cardiovasc Imaging. 2022;15:1091–1102. doi:doi:10.1016/j.jcmg.2021.04.030. [PubMed: 34274267]
7. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. IEEE Int Conf Comput Vis. 2017;1:618–26.
8. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Bossuyt P, et al. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17:230. doi:10.1186/s12916-019-1466-7. [PubMed: 31842878]
9. Van Calster B, Vickers AJ. Calibration of risk prediction models: impact on decision-analytic performance. Med Decis Making. 2015;35:162–9. doi:10.1177/0272989X14547233. [PubMed: 25155798]
10. Slomka PJ, Betancur J, Liang JX, Otaki Y, Hu L, Sharir T, et al. Rationale and design of the REgistry of Fast Myocardial Perfusion Imaging with NExt generation SPECT (REFINE SPECT). J Nucl Cardiol. 2020;27:1010–1021. [PubMed: 29923104]
11. Hu L-H, Sharir T, Miller RJH, Einstein AJ, Fish MB, Ruddy TD, et al. Upper reference limits of transient ischemic dilation ratio for different protocols on new-generation cadmium zinc telluride cameras: A report from REFINE SPECT registry. J Nucl Cardiol. 2020;27:1180–9. doi:10.1007/s12350-019-01730-y. [PubMed: 31087268]
12. Diamond GA, Forrester JS. Analysis of probability as an aid in the clinical diagnosis of coronary-artery disease. N Engl J Med. 1979;300:1350–8. [PubMed: 440357]
13. Miller RJ, Klein E, Gransar H, Slomka PJ, Friedman JD, Hayes S, et al. Prognostic significance of previous myocardial infarction and previous revascularization in patients undergoing SPECT MPI. Int J Cardiol. 2020;313:9–15. [PubMed: 32349938]
14. Gambhir SS, Berman DS, Ziffer J, Nagler M, Sandler M, Patton J, et al. A novel high-sensitivity rapid-acquisition single-photon cardiac imaging camera. J Nucl Med. 2009;50:635–43. doi:10.2967/jnumed.108.060020. [PubMed: 19339672]
15. Dorbala S, Ananthasubramaniam K, Armstrong IS, Chareonthaitawee P, DePuey EG, Einstein AJ, et al. Single Photon Emission Computed Tomography (SPECT) Myocardial Perfusion

Imaging Guidelines: Instrumentation, Acquisition, Processing, and Interpretation. J Nucl Cardiol. 2018;25:1784–846. doi:10.1007/s12350-018-1283-y. [PubMed: 29802599]

16. Cerqueira MD, Weissman NJ, Dilsizian V, Jacobs AK, Kaul S, Laskey WK, et al. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart. A statement for healthcare professionals from the Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association. Int J Cardiovasc Imaging. 2002;18:539–42. [PubMed: 12135124]

17. Slomka PJ, Nishina H, Berman DS, Akincioglu C, Abidov A, Friedman JD, et al. Automated quantification of myocardial perfusion SPECT using simplified normal limits. J Nucl Cardiol. 2005;12:66–77. doi:10.1016/j.nuclcard.2004.10.006. [PubMed: 15682367]

18. Arsanjani R, Xu Y, Hayes SW, Fish M, Lemley M Jr., Gerlach J, et al. Comparison of fully automated computer analysis and visual scoring for detection of coronary artery disease from myocardial perfusion SPECT in a large population. J Nucl Med. 2013;54:221–8. doi:10.2967/jnumed.112.108969. [PubMed: 23315665]

19. Ferro C Comparing Probabilistic Forecasting Systems with the Brier Score. Weather Forecast. 2007;22:1076–88. doi:10.1175/WAF1034.1.

20. Molinaro AM, Simon R, Pfeiffer RM. Prediction error estimation: a comparison of resampling methods. Bioinform. 2005;21:3301–7. doi:10.1093/bioinformatics/bti499.

21. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143:29–36. doi:10.1148/radiology.143.1.7063747. [PubMed: 7063747]

22. Rengasamy D, Jafari M, Rothwell B, Chen X, Figueredo GP. Deep Learning with Dynamically Weighted Loss Function for Sensor-Based Prognostics and Health Management. Sensors. 2020;20:723. doi:10.3390/s20030723. [PubMed: 32012944]

23. Shorten C, Khoshgoftaar TM. A survey on Image Data Augmentation for Deep Learning. J Big Data. 2019;6:60. doi:10.1186/s40537-019-0197-0.

24. Rozanski A, Miller RJH, Han D, Gransar H, Slomka P, Dey D, et al. The prevalence and predictors of inducible myocardial ischemia among patients referred for radionuclide stress testing. J Nucl Cardiol. 2021. Epub ahead of print. doi:10.1007/s12350-021-02797-2.

25. Manfrini O, Yoon J, Schaar Mvd, Kedev S, Vavlukis M, Stankovic G, et al. Sex Differences in Modifiable Risk Factors and Severity of Coronary Artery Disease. JAHA. 2020;9:e017235. doi:doi:10.1161/JAHA.120.017235. [PubMed: 32981423]

26. Iskandar A, Limone B, Parker MW, Perugini A, Kim H, Jones C, et al. Gender differences in the diagnostic accuracy of SPECT myocardial perfusion imaging: a bivariate meta-analysis. J Nucl Cardiol. 2013;20:53–63. doi:10.1007/s12350-012-9646-2. [PubMed: 23149886]

27. Apostolopoulos ID, Papathanasiou ND, Spyridonidis T, Apostolopoulos DJ. Automatic characterization of myocardial perfusion imaging polar maps employing deep learning and data augmentation. Hell J Nucl Med. 2020;23:125–32. doi:10.1967/s002449912101. [PubMed: 32716403]

28. Miller RJH, Kuronuma K, Singh A, Otaki Y, Hayes S, Chareonthaitawee P, et al. Explainable Deep Learning Improves Physician Interpretation of Myocardial Perfusion Imaging. J Nucl Med. 2022:jnumed.121.263686. doi:10.2967/jnumed.121.263686.

29. Gimelli A, Bottai M, Quaranta A, Giorgetti A, Genovesi D, Marzullo P. Gender differences in the evaluation of coronary artery disease with a cadmium-zinc telluride camera. Eur J Nucl Med Mol Imaging. 2013;40:1542–8. doi:10.1007/s00259-013-2449-0. [PubMed: 23703458]

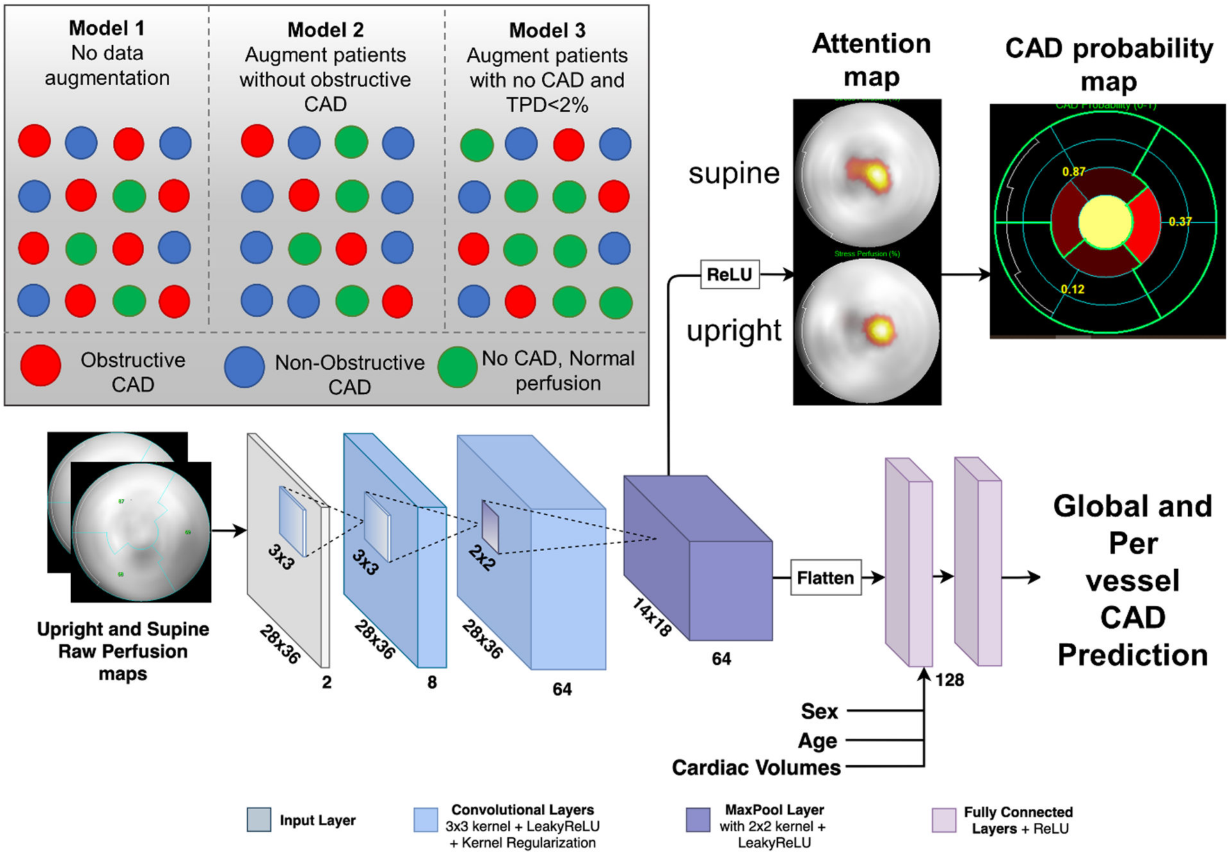
Model 1
No data augmentation
Model 2
Augment patients without obstructive CAD
Model 3
Augment patients with no CAD and TPD<2%
Obstructive CAD
Non-Obstructive CAD
No CAD, Normal perfusion
Attention map
CAD probability map
supine
upright
ReLU
0.87
0.37
0.12
Upright and Supine Raw Perfusion maps
3x3
3x3
2x2
28x36
28x36
28x36
14x18
2
8
64
64
Flatten
128
Global and Per vessel CAD Prediction
Sex
Age
Cardiac Volumes
Input Layer
Convolutional Layers 3x3 kernel + LeakyReLU + Kernel Regularization
MaxPool Layer with 2x2 kernel + LeakyReLU
Fully Connected Layers + ReLU

Fig. 1.

Deep Learning Architecture. The raw polar maps are input to the network without the use of pre-defined coronary territories or any assumed subdivision. Patient sex, age, BMI, and cardiac volumes information is added to the final feature vector. The attention map highlights regions contributing most to the DL score for a given patient, while the CAD probability map shows per-vessel probability of obstructive CAD. CAD=coronary artery disease, Grad-CAM=Gradient-weighted Class Activation Mapping, BMI=body mass index.

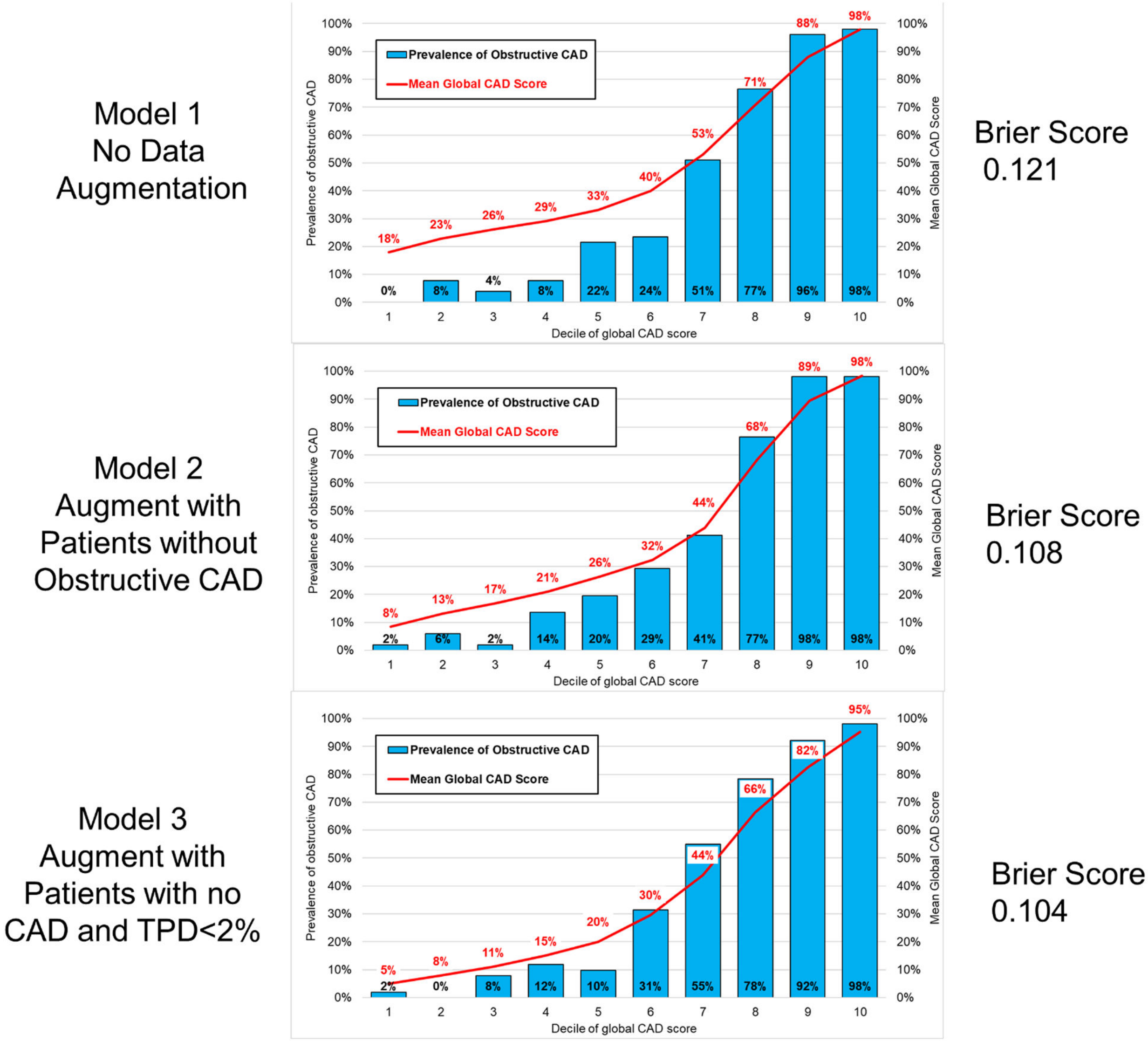

Fig. 2.

Calibration graphs in the external patient population showing predicted probability of CAD compared to actual prevalence of CAD. Calibration was good for all models but was significantly higher for Model 2 and Model 3 compared to Model 1. CAD – coronary artery disease.

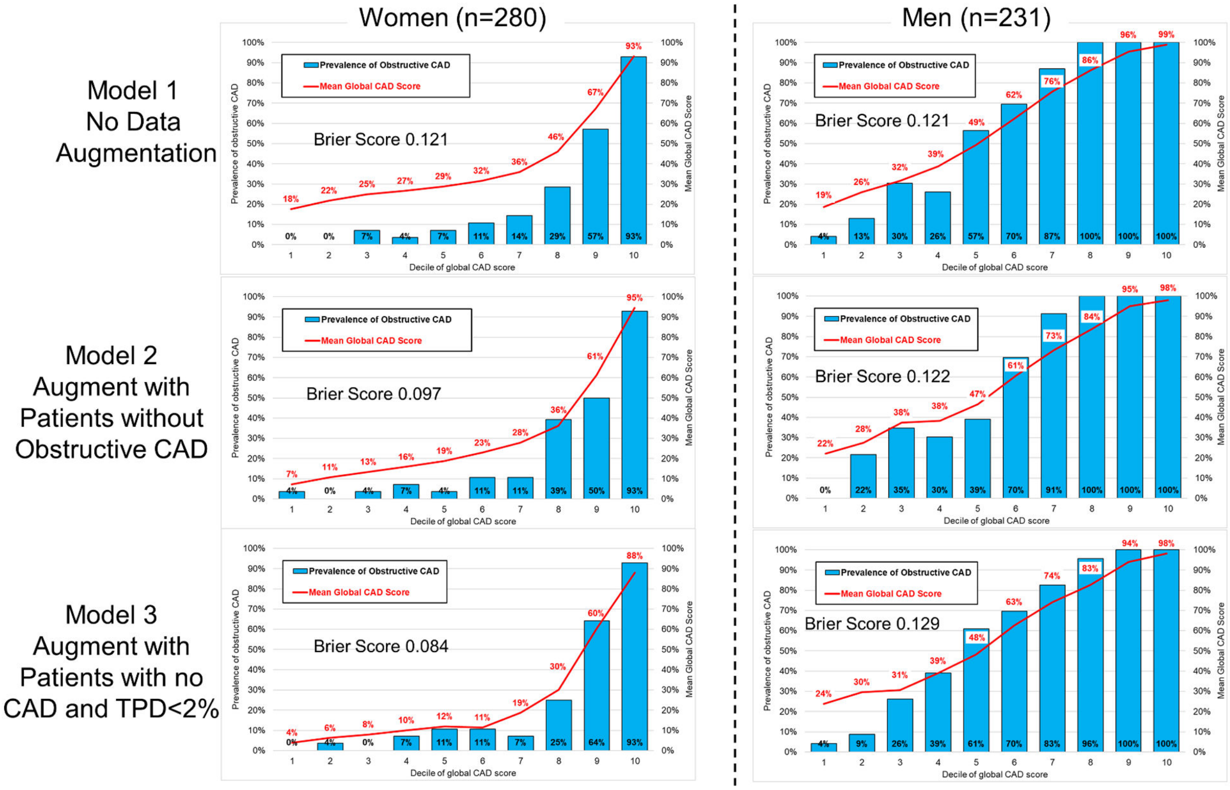

Fig 3.

Calibration graphs in women and men showing predicted probability of CAD compared to actual prevalence of CAD. Calibration in women was best for Model 3, with both Model 2 and Model 3 having significantly better calibration than Model 1 (both p<0.001). Calibration was similar for all models in men.

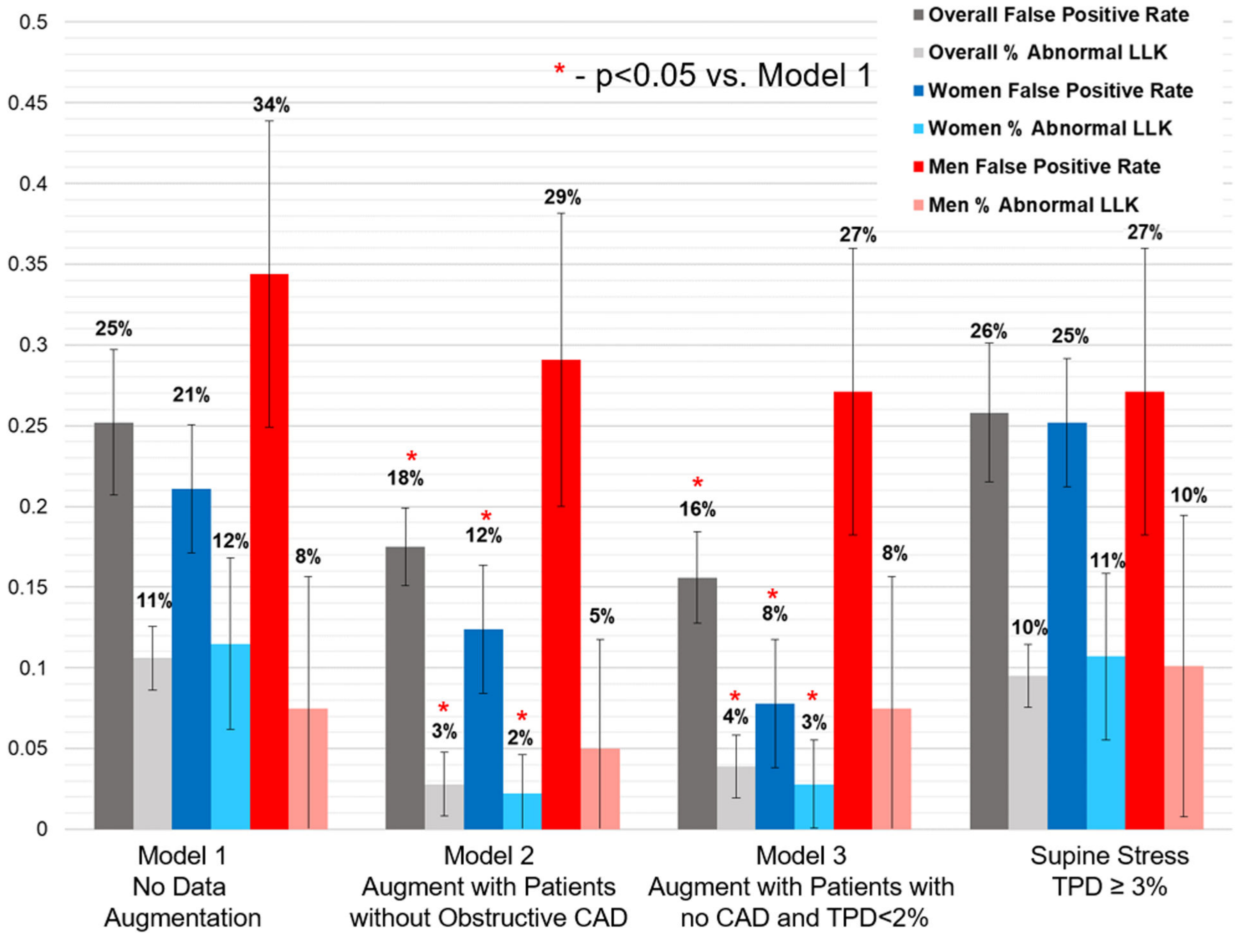

Fig. 4.

False positive rates and rates of abnormality in patients with low likelihood (LLK) of coronary artery disease (CAD). Model 1 was trained with no data augmentation, but with a weighted loss function. Training for model 2 was augmented with additional patients without obstructive CAD. Training for model 3 was augmented with additional patients without CAD and with total perfusion deficit < 2%. False positive rates and abnormality in LLK were significanly lower for Model 2 and 3 compared to model 1 for the overall population and in women (* p<0.05), but not in men. TPD – total perfusion deficit.

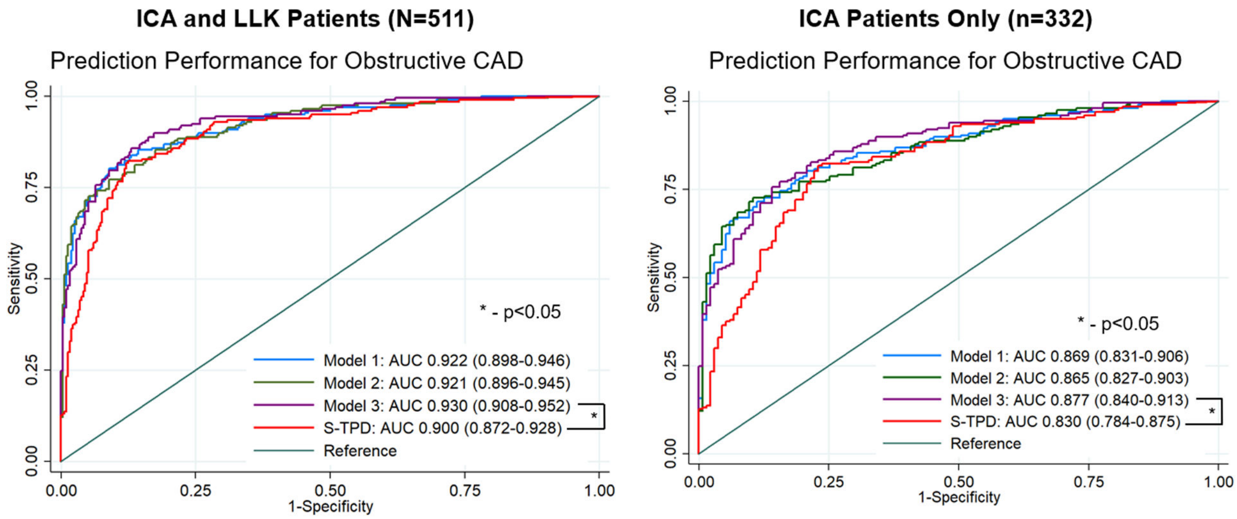

Fig. 5.

Diagnostic accuracy for obstructive coronary artery disease in the external population using different training augmentation methods. Left panel shows the entire population (patients undergoing invasive coronary angiography [ICA] and low-likelihood (LLK) patients). Right panel shows only patients who underwent ICA. Model 1 was trained without data augmentation, but using a weighted loss for training. Model 2 augments training with patients without obstructive CAD and model 3 augments training with patients without CAD and total perfusion deficit <2%. Only Model 3 had significantly higher prediction performance compared to supine stress total perfusion deficit (S-TPD). AUC – area under the receiver operating characteristic curve.

Table 1.

Training and Testing Population Characteristics

	Training Population (n=828)	Testing Population (n=511)	p-value
Age, years	64.0 ±11.6	60.8 ± 12.2	<0.001
BMI, kg/m^2	28.7 [25.4 – 32.3]	29.9 [25.9 – 34.0]	<0.001
Females	274 (33.1)	280 (54.8)	<0.001
Hypertension	598 (72.2)	318 (62.2)	<0.001
Diabetes	245 (29.6)	106 (20.7)	<0.001
Dyslipidemia	560 (67.6)	290 (56.8)	<0.001
Smoking	108 (13.0)	151 (29.6)	<0.001
Typical angina	113 (13.7)	117 (22.9)	<0.001
Atypical angina	337 (40.7)	103 (20.2)	<0.001
Asymptomatic	324 (39.1)	149 (29.2)	<0.001
Exercise stress	419 (50.6)	282 (55.2)	0.092
U-TPD (%)	7.3 [3.1 – 14.7)	2.2 [0.6 – 9.0)	<0.001
S-TPD (%)	8.4 [3.7 – 16.3]	3.1 [0.9 – 10.7)	<0.001
ICA – obstructive CAD	521 (62.9)	197 (38.5)	<0.001
ICA – no obstructive CAD	307 (37.1)	135 (26.4)	<0.001
Low likelihood of CAD	0 (0)	179 (35)	<0.001

Categorical values are expressed as n (%). Continuous value is expressed as mean ± SD or median [interquartile range]. Patients who underwent invasive coronary angiography (ICA) were classified as having obstructive coronary artery disease (CAD) if there was any stenosis ≥70% or left main stenosis ≥ 50%. BMI - body mass index, CAD – coronary artery disease, TPD - total perfusion deficit.

Table 2.

Angiographic Characteristics

	Training Population (n=828)	Testing Population (n=332)	p-value
1 vessel disease	241 (29.1)	80 (24.1)	0.095
2 vessel disease	171 (20.7)	69 (20.8)	1.000
3 vessel disease	109 (13.2)	48 (14.5)	0.569
Left main disease	50 (6.7)	26 (7.9)	0.294
LAD disease	371 (44.8)	138 (41.6)	0.327
LCx disease	273 (33.0)	111 (33.4)	0.890
RCA disease	266 (32.1)	113 (34.0)	0.534

Value is expressed as n (%). LAD - left anterior descending artery, LCx - left circumflex coronary artery, RCA - right coronary artery.