

1 **Explainable Deep Learning Improves Physician Interpretation of Myocardial Perfusion**
2 **Imaging**

3 Robert JH Miller, MD^{1,2}, Keiichiro Kuronuma, MD, PhD^{1,3}, Ananya Singh, MSc¹, Yuka Otaki,
4 MD, PhD¹, Sean Hayes, MD¹, Panithaya Chareonthaitawee, MD⁴, Paul Kavanagh, MS¹, Tejas
5 Parekh, BSc¹, Balaji K. Tamarappoo, MD, PhD¹, Tali Sharir, MD⁵, Andrew J. Einstein, MD,
6 PhD⁶, Mathews B. Fish, MD⁷, Terrence D. Ruddy, MD⁸, Philipp A. Kaufmann, MD⁹, Albert J.
7 Sinusas, MD¹⁰, Edward J. Miller, MD PhD¹⁰, Timothy M. Bateman, MD¹¹, Sharmila Dorbala,
8 MD, MPH¹², Marcelo Di Carli, MD¹², Sebastien Cadet, MSc¹, Joanna X. Liang, BA¹, Damini
9 Dey, PhD¹, Daniel S. Berman, MD¹, Piotr J. Slomka, PhD¹

10 1 Department of Medicine (Division of Artificial Intelligence), Imaging, and Biomedical Sciences,
11 Cedars-Sinai Medical Center, Los Angeles, CA, USA

12 2 Department of Cardiac Sciences, University of Calgary and Libin Cardiovascular Institute,
13 Calgary AB, Canada

14 3 Department of Cardiology, Nihon University, Tokyo, Japan

15 4 Department of Cardiovascular Medicine, Mayo Clinic, Rochester MN

16 5 Department of Nuclear Cardiology, Assuta Medical Centers, Tel Aviv, Israel and Ben Gurion
17 University of the Negev, Beer Sheva, Israel

18 6 Division of Cardiology, Department of Medicine, and Department of Radiology, Columbia
19 University Irving Medical Center and New York-Presbyterian Hospital, New York, NY, USA.

20 7 Oregon Heart and Vascular Institute, Sacred Heart Medical Center, Springfield, OR, USA.

21 8 Division of Cardiology, University of Ottawa Heart Institute, Ottawa, ON, Canada

22 9 Department of Nuclear Medicine, Cardiac Imaging, University Hospital Zurich, Zurich,
23 Switzerland

24 10 Section of Cardiovascular Medicine, Department of Internal Medicine, Yale University School
25 of Medicine, New Haven, CT, USA.

26 11 Cardiovascular Imaging Technologies LLC, Kansas City, MO, USA

27 12 Department of Radiology, Division of Nuclear Medicine and Molecular Imaging, Brigham and
28 Women's Hospital, Boston, MA, USA

29

30 **Address for Correspondence:**

31 **Piotr Slomka, PhD**

32 Director of Innovation in Imaging, Cedars-Sinai

33 Professor of Medicine and Cardiology

34 Division of Artificial Intelligence in Medicine, Cedars-Sinai

35 8700 Beverly Blvd., Ste. Metro 203

36 Los Angeles, California 90048

37 Phone: 310-423-4348 Fax: 310-423-0173

38 Email: Piotr.Slomka@cshs.org

39

- 1 **First Author:**
- 2 Robert JH Miller, MD
- 3 Clinical assistant professor, University of Calgary
- 4 Department of Cardiac Sciences
- 5 GAA 08, Health research innovation center
- 6 3230 Hospital Drive NW, Calgary, AC T2N 4Z6

- 7 **Word count:** 4797 words (including all components)

- 8 **Short Title:** Explainable AI Improves MPI Accuracy

- 9

1 **ABSTRACT**

2 **Rationale:**

3 Artificial intelligence may improve accuracy of myocardial perfusion imaging (MPI) but will
4 likely be implemented as an aid to physician interpretation rather than an autonomous tool. Deep
5 learning (DL) has high standalone diagnostic accuracy for obstructive coronary artery disease
6 (CAD), but its influence on physician interpretation is unknown. We assessed whether access to
7 explainable DL predictions improves physician interpretation of MPI.

8 **Methods:**

9 We selected a representative cohort of patients who underwent MPI with reference invasive
10 coronary angiography. Obstructive CAD, defined as stenosis $\geq 50\%$ in the left main artery or $\geq 70\%$
11 in other coronary segments, was present in half of patients. We utilized an explainable DL model
12 (CAD-DL), which was previously developed in a separate population from different sites. Three
13 physicians interpreted studies first with clinical history, stress, and quantitative perfusion, then
14 with all the data plus the DL results. Diagnostic accuracy was assessed using area under the
15 receiver-operating characteristic curve (AUC).

16 **Results:**

17 In total, 240 patients were included with median age 65 (IQR 58 – 73). The diagnostic accuracy
18 of physician interpretation with CAD-DL (AUC 0.779) was significantly higher compared to
19 physician interpretation without CAD-DL (AUC 0.747, $p=0.003$) and stress total perfusion deficit
20 (AUC 0.718, $p<0.001$). With matched specificity, CAD-DL had higher sensitivity when operating
21 autonomously compared to readers without DL results ($p<0.001$), but not compared to readers
22 interpreting with DL results ($p=0.122$). All readers had numerically higher accuracy with the use
23 of CAD-DL, with AUC improvement 0.02 to 0.05, and interpretation with DL resulted in overall
24 net reclassification improvement of 17.5% (95% CI 9.8% – 24.7%, $p<0.001$).

25 **Conclusion:**

26 Explainable DL predictions lead to meaningful improvements in physician interpretation;
27 however, the improvement varied across the readers reflecting the acceptance of this new
28 technology. This technique could be implemented as an aid to physician diagnosis, improving
29 the diagnostic accuracy of MPI.

30

31 **KEYWORDS:** Artificial intelligence, deep learning, implementation

1 **ABBREVIATIONS**

2 AI: artificial intelligence

3 AUC: area under the ROC curve

4 CAD: coronary artery disease

5 LAD: left anterior descending coronary artery

6 LCx: left circumflex coronary artery

7 MPI: myocardial perfusion imaging

8 RCA: right coronary artery

9 SPECT: single photon emission computed tomography

10 SSS: summed stress score

11 TPD: total perfusion deficit

12

13

14

1 INTRODUCTION

2 Coronary artery disease (CAD) is a major cause of death in the United States (1), making
3 accurate diagnosis critically important. Myocardial perfusion imaging (MPI) is frequently used
4 to diagnose obstructive CAD(2), predict cardiovascular risk(3), or guide treatment
5 decisions(4,5). Artificial intelligence (AI) may be able to improve the diagnostic(6-9), and
6 prognostic accuracy of myocardial perfusion imaging (MPI)(10,11); however, it is not feasible or
7 desirable to implement it without physician oversight (12-15). As a result, for the foreseeable
8 future, AI will most likely be implemented as an aid to physician interpretation rather than
9 operating autonomously (12-15).

10 We recently demonstrated an approach for automated interpretation of MPI by a general
11 purpose deep-learning (CAD-DL) model which incorporated two methods to explain predictions
12 to physicians(9). These methods for explainability allow physicians to ensure AI findings are
13 clinically relevant and potentially identify errors in either AI or physician interpretations.
14 Explainable AI is critical to overcoming the “black-box” perception of AI (16,17), and is
15 recognized as an important advancement in a recent AI best practice statement (12). In our initial
16 study, autonomous CAD-DL score had higher diagnostic accuracy for obstructive CAD
17 compared to expert visual interpretation and quantitative assessment of myocardial perfusion(9).
18 However, it remains to be shown if explainable DL predictions, utilized as an aid during clinical
19 interpretation, can improve physician interpretation of MPI, which is the likely method for
20 clinical implementation (12).

21 Accordingly, we performed a prospective study to assess the potential improvement in
22 accuracy of physician interpretation of SPECT MPI that could be achieved by utilizing the

1 explainable DL model as an aid during clinical interpretation, using an external population from
2 sites not used for model training.

3

4 **MATERIALS AND METHODS**

5 Study Population

6 We included 240 patients who underwent single photon emission computed tomography
7 (SPECT) MPI with reference coronary angiography within 180 days. Patients were included
8 from two sites, Columbia University (n=125) and Cardiovascular Imaging Technologies LLC
9 (n=115). No data from these sites were used in the development of CAD-DL. Patients were
10 randomly selected to include a representative patient population with prevalence of obstructive
11 CAD of 50%. Patients underwent imaging with either conventional Anger camera systems
12 (n=80) or solid-state camera systems (n=160). Patients underwent stress-rest (n=158), rest-stress
13 (n=61) or stress-only (n=21) imaging protocols with either exercise stress with a symptom-
14 limited Bruce protocol (n=98) or pharmacologic stress with adenosine (n=31) or regadenoson
15 (n=111). Patients underwent either a standard single-isotope (n=205) or dual-isotope (n=35)
16 SPECT MPI protocol as previously described (2,18). For comparison, CAD-DL was trained in a
17 population imaged with 58% solid-state camera systems and 42% conventional camera systems,
18 with 69% of patients undergoing pharmacologic stress. The institutional review board at each
19 site approved this study with either signed informed consent or waiver of informed consent.

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

1 sharing agreements and institutional review board protocols, the data from this manuscript will
2 be shared upon written request.

3 Details regarding invasive coronary angiography and image quantification are available
4 in the supplement (19,20).

5 Deep Learning Model Architecture

6 The architecture for CAD-DL has previously been described in detail (9). In brief, CAD-
7 DL was trained using raw polar maps, pre-processed using Z-normalization (mean of 0 and
8 standard deviation of 1), of myocardial perfusion, wall motion, and thickening(21), as well as
9 age, sex, left ventricular end-systolic volume and end-diastolic volumes which were all obtained
10 automatically from image data. CAD-DL was trained in a separate population of 3,578 patients,
11 with obstructive CAD present in 63% of patients. CAD-DL was implemented using Python 3.7.3
12 and Tensorflow 1.14. The training was performed using Titan RTX graphics card (Nvidia, Santa
13 Clara, CA). The model was trained using 5-fold cross-validation in the previous study, which did
14 not include any data from the 2 sites in the present study.

15 CAD-DL includes two methods to explain predictions: attention maps and probability
16 maps. Attention maps utilize Gradient-weighted Class Activation Mapping (22) to highlight
17 myocardial regions which contributed most to the prediction. Using these attention maps, each
18 segment from the standard 17-segment American Heart Association model can be assigned to 5
19 categories to generate the segmental CAD probability map. The CAD probability map marks the
20 degree to which the segments contribute the CAD-DL prediction, to give further insight for the
21 clinical reader.

22

1 Physician Interpretation

2 Three physicians with a range of clinical experience (2 years to >20 years in clinical
3 practice) interpreted all cases in duplicate. Initially, readers interpreted myocardial perfusion
4 images in the conventional manner and were supplied by the following: age, sex, body mass
5 index (BMI), past medical history, test indication, electrocardiographic stress response, and
6 images. Readers had access to all image data sets including stress and rest as well as supine and
7 upright when available and gated imaging for all studies(23). Raw data was available for quality
8 control and standard quantitative measures of function and perfusion were available for all
9 studies. Readers interpreted the overall study with a 5-point scale (normal, probably normal,
10 equivocal, probably abnormal, and definitely abnormal). Readers also interpreted studies with
11 semi-quantitative scoring to generate summed stress score (SSS), summed rest score (SRS) and
12 summed difference score (SDS) using the standard 17-segment model.

13 After the initial interpretation, readers interpreted the study in conjunction with CAD-DL.
14 The clinical prototype of CAD-DL developed previously, with attention and probability maps,
15 was integrated within standard clinical nuclear cardiology software (QPS/QGS Cedars-Sinai
16 Medical Center, Los Angeles, CA). Physicians repeated the interpretation process when
17 informed by CAD-DL results. This process was designed to simulate the use of CAD-DL in
18 clinical practice, where it would be incorporated as an expert second reader. Physicians were
19 trained on how to generate CAD-DL results but were not given specific thresholds for CAD-DL
20 global score, attention map, or CAD-DL probability scores to apply in their new interpretation.
21 No specific instructions were given for adjusting reader interpretations based on CAD-DL
22 results. This was done purposefully so the results would reflect clinical practice whether other

1 factors (e.g., reader confidence in original interpretation, belief in AI interpretation) would
2 influence the degree to which readers alter their interpretation.

3 Statistical Analysis

4 In the primary analysis, the diagnostic performance of SSS without DL, SSS with DL,
5 and stress TPD was evaluated using the analysis and pairwise comparisons of the areas under the
6 receiver operating characteristic (ROC) curve (AUC) according to DeLong et al.(24) in order to
7 allow comparisons with automated perfusion assessment. However, in order to more fully assess
8 the impact of DL predictions on reader diagnostic accuracy we performed an analysis to account
9 for the multiple reader, multiple case (MRMC) design which accounts for variation related to
10 case variation, reader certainty, and reader skill. In this analysis, a random-effects ROC analysis
11 was used to compare the reader-averaged nonparametric AUC with and without access to DL
12 predictions as previously described(25,26). Additional details are available in the supplement.

13

14 RESULTS

15 Clinical Characteristics and Angiographic Characteristics

16 In total, 240 patients were included in this study with median age 65 (IQR 58 – 73) and
17 156 (65.0%) male. Invasive angiography was performed at a median of 11 days (IQR 3 – 27
18 days) post SPECT MPI. Obstructive CAD was present in 120 (50.0%) patients including: 11
19 patients with left main, 84 with LAD, 55 with LCx and 63 with RCA disease. Characteristics in
20 patients with and without obstructive CAD are shown in Table 1. Characteristics stratified by
21 camera type are shown in Supplemental Table 1. Median age was similar in patients imaged with
22 a conventional or solid-state camera system (median 66 vs 65, p=0.858).

1 Per Patient Diagnostic Accuracy

2 ROC curves for identification of obstructive CAD based on stress perfusion assessment
3 (SSS) are shown in **Figure 1**. The AUC of physician interpretation with DL (AUC 0.779, 95% CI
4 0.738 – 0.850) was significantly higher compared to both physician interpretation without DL
5 (AUC 0.747, 95% CI 0.685 – 0.809, $p=0.003$) and stress TPD (AUC 0.718, 95% CI 0.653 –
6 0.782, $p<0.001$). The diagnostic accuracy of physician interpretation with DL was similar to
7 CAD-DL operating autonomously (AUC 0.793, 95% CI 0.736-0.849, $p=0.536$).

8 Diagnostic accuracy for each reader separately is shown in **Supplemental Figure 1**. There
9 was a trend towards improvement in accuracy for two readers (Reader 2 AUC 0.750 vs 0.730,
10 $p=0.115$ and Reader 3 AUC 0.751 vs. 0.733, $p=0.068$). Reader 1 demonstrated significantly
11 improved accuracy with access to CAD-DL predictions (AUC 0.805 vs. 0.756, $p=0.005$).
12 Readers 1 and 2 were less experienced (4 and 2 years, respectively) compared to Reader 3 (>20
13 years). For comparison, the AUC of DL operating autonomously was 0.793 (95% CI 0.736-
14 0.849). In the MRMC analysis, reader accuracy was also significantly improved with access to
15 CAD-DL predictions (AUC 0.769 vs. 0.740, $p=0.019$).

16 **Figure 2** shows reader sensitivity and specificity, using a threshold of $SSS>3$, with and
17 without DL. All test characteristics numerically improved when readers had access to
18 explainable DL results, with improvement in both sensitivity and specificity when considering
19 all readers together (both $p<0.01$). With matched specificity, CAD-DL had higher sensitivity
20 when operating autonomously compared to readers without access to explainable DL results
21 ($p<0.001$). However, this difference was not significant when readers had access to DL
22 predictions during interpretation ($p=0.122$).

1 Our study was powered to detect a difference in accuracy for the overall population.
2 However, there was a trend towards improved physician diagnostic performance with DL
3 compared to without DL for both camera systems (solid-state AUC 0.799 vs 0.774, $p=0.095$;
4 conventional AUC 0.740 vs 0.691, $p=0.014$) across imaging protocols (stress-rest AUC 0.775 vs
5 0.738, $p=0.009$; rest-stress AUC 0.710 vs 0.690, $p=0.464$; and stress only AUC 0.891 vs 0.881,
6 $p=0.665$), in men (AUC 0.792 vs 0.772, $p=0.096$) and women (AUC 0.714 vs 0.658,) $p=0.028$,
7 and in patients undergoing exercise (AUC 0.816 vs 0.795, $p=0.250$) or pharmacologic stress
8 (AUC 0.728 vs 0.692, $p=0.020$). Reader interpretation of ischemia with DL (based on SDS) also
9 demonstrated significantly higher AUC compared to reader interpretation without DL or
10 ischemic TPD (Figure 3).

11 Lastly, reader diagnosis, using a 5-point scale, with DL also demonstrated significantly
12 higher accuracy compared to reader diagnosis without DL (Supplemental Figure 2). The re-
13 classification of patients according to reader diagnosis for the three readers is shown in Table 2.
14 There was an overall net reclassification improvement of 17.2% (95% CI 9.2% – 24.4%,
15 $p<0.001$), with improved classification of patients with CAD of 6.1% (95% CI 1.4% - 10.3%)
16 and patients without CAD of 11.1% (95% CI 4.8% - 16.8%). When interpreting with DL
17 compared to without DL, there was no difference in the proportion of cases interpreted as
18 equivocal (9.6% vs.8.6%, $p=0.529$). Similarly, there was no difference in the proportion of cases
19 with CAD interpreted as definitely abnormal (59% vs 58%, $p=0.803$) or patients without CAD
20 interpreted as definitely normal (26% vs 23%, $p=0.464$).

21 One case where all 3 physicians increased their segmental scores in a patient with
22 obstructive CAD is shown in Supplemental Figure 3. One case where all 3 physicians decreased
23 their segmental scores in a patient without obstructive CAD is shown in Supplemental Figure 4.

1 A description of all cases in which all readers increased or decreased their scores is available in
2 [Supplemental Table 2](#). An example of a case with high CAD-DL score which was not
3 consistently scored as abnormal is shown in [Supplemental Figure 5](#).

4 Per Vessel Diagnostic Accuracy

5 AUCs for identification of obstructive CAD for each vessel is shown in [Supplemental](#)
6 [Figure 6](#). Reader diagnostic accuracy with DL (AUC 0.723, 95% CI 0.652 – 0.793) was
7 significantly better compared to accuracy without DL (0.697, 95% CI 0.626 – 0.768, $p=0.041$)
8 for LAD disease. Reader interpretation with DL had higher AUC compared to reader
9 interpretation without DL and stress TPD for left anterior descending disease ($p=0.041$ and
10 $p=0.022$ respectively). Reader interpretation with DL was not significantly higher compared to
11 reader interpretation without DL for left circumflex (AUC 0.727 vs. 0.715, $p=0.529$) or right
12 coronary artery disease (AUC 0.776 vs 0.779, $p=0.597$). Reclassification according to vascular
13 territory is shown in [Supplemental Table 3](#).

14

15 **DISCUSSION**

16 We performed a prospective study in an external population to determine the potential
17 influence of utilizing an explainable DL model as an interpretation aid on physician diagnostic
18 accuracy. We demonstrated that overall physician interpretation significantly improved by
19 utilizing the DL predictions compared to the same physicians interpreting without DL.
20 Additionally, with the aid of DL physician interpretation had higher diagnostic accuracy
21 compared to quantitative assessment of perfusion. There was a trend towards higher diagnostic
22 performance for every reader, and results were consistent across camera systems, imaging

1 protocols and patient subgroups. There was some heterogeneity in improvement in physician
2 diagnostic performance; however, there was more uniformity in sensitivity and specificity across
3 readers when interpreting with DL results. All of these advancements were demonstrated despite
4 the relative novelty of the DL tool and lack of physician experience with using the new DL
5 module. Our results suggest that implementing DL as an aid to physician interpretation could
6 significantly improve diagnostic accuracy of MPI.

7 Several studies have previously demonstrated that AI algorithms can be used to achieve
8 high diagnostic accuracy of SPECT MPI. Arsanjani et al. demonstrated that a support vector
9 machines model, improved diagnostic accuracy for obstructive CAD compared to quantitative
10 assessment of perfusion with TPD(27). Betancur et al. demonstrated that a different DL model
11 improved detection of obstructive CAD compared to quantitation of perfusion with TPD on both
12 a regional and per-patient basis, in a study that included 1638 patients from 9 centers.(6) With
13 matched specificity, DL improved the per-vessel sensitivity to 70% compared to 64% with TPD
14 ($p<0.01$). (6) Subsequently our group demonstrated that a DL algorithm utilizing both upright and
15 supine imaging data improved diagnostic accuracy compared to combined TPD analysis.(7)
16 Spier et al. demonstrated that a convolutional neural network could classify stress polar maps as
17 normal or abnormal with excellent agreement with expert interpretation (91.1%) (28). More
18 recently we demonstrated that the current model, CAD-DL, had higher diagnostic accuracy for
19 obstructive CAD compared to physician interpretation or quantitative assessment of
20 perfusion(9). However, all of these studies demonstrated only standalone performance, without
21 physician oversight, which is not practical clinically.

22 In the current study we addressed an important knowledge gap by determining to what
23 extent access to explainable DL predictions could influence or improve physician interpretation.

1 We demonstrated, in an external population suggesting broader generalizability (12), that overall
2 physician interpretation significantly improved with the aid of CAD-DL compared to the same
3 physicians interpreting without CAD-DL. All readers demonstrated numerically higher AUC
4 with the use of DL, with one reader improving significantly. Importantly, readers had access to
5 the same clinical and imaging information (including quantitative results) during each
6 interpretation, with the only difference being the AI predictions. Overall reader sensitivity and
7 specificity both improved, achieving similar results to CAD-DL operating autonomously. Using
8 CAD-DL as an aid also significantly improved classification of the overall population, patients
9 with CAD, and patients without CAD. Per-vessel diagnostic accuracy was significantly higher
10 for the left anterior descending and similar in other vascular territories. There was also net
11 improvement in classification across all territories. This result was obtained using a patient
12 population separate from the original population used for training, with different population
13 characteristics and prevalence of obstructive CAD suggesting good generalizability of the
14 results. Importantly, the DL model was incorporated into standard clinical interpretation software
15 which was used by all readers and generated results in <10 seconds. However, we did not
16 measure the average interpretation time for readers when interpreting with and without the
17 explainable DL predictions which may be an important consideration for clinical
18 implementation. Overall, our findings suggest that our model could be implemented into clinical
19 practice as an aid to physician interpretation in order to improve the diagnostic accuracy of
20 SPECT MPI.

21 While the potential benefits of AI for improving diagnostic accuracy are becoming
22 readily apparent, practical questions about clinical implementation have remained. One step
23 towards clinical implementation is the development of models capable of explaining results to

1 the physician. The CAD-DL model user interface includes two methods to explain predictions to
2 the clinicians with the attention and probability maps. In order to replicate future clinical
3 implementation, we instructed physicians on how to access predictions but did not explicitly
4 instruct them on how to incorporate this information. This approach to utilizing DL predictions
5 mirrors future clinical use where factors such as physician experience, confidence in original
6 interpretation, belief in AI, and anchoring bias would influence thresholds for changing
7 interpretation and magnitude of change (29). As a result, and as was seen in our study, the degree
8 to which CAD-DL influences interpretation varies between physicians (Figure 3). Although there
9 was no clear relationship between reader experience and improvement in accuracy in our study,
10 it is likely that less experienced readers would derive more benefit. One recent study suggested
11 that a deep-learning algorithm could be used by novice physicians to attain similar interpretation
12 of myocardial perfusion to experts(30). Additionally, further improvements may be possible if
13 physicians develop more experience with incorporating the DL predictions. In spite of this
14 variation, access to DL results led to more uniform sensitivity and specificity across readers.
15 Additionally, as physician experience with DL increases, diagnostic accuracy may improve
16 further. Incorporating AI as an aid to physician interpretation avoids potential medicolegal issues
17 related to using these technologies because the final responsibility for interpretation still lies with
18 the physician.

19 Our study has a few important limitations. The explainable DL model incorporates several
20 methods of explaining results (attention maps, probability maps, per-vessel probabilities) which
21 are presented simultaneously in the clinical module. We did not elucidate which aspect of the
22 explainable DL predictions lead to improved accuracy since they are designed to be interpreted
23 collectively. The per-vessel probabilities convey the likelihood of CAD to physicians while the

1 attention and probability maps direct the physician's attention to allow them to validate the
2 accuracy of those predictions. Additionally, we did not measure subjective changes in reader
3 confidence but did not identify a change in proportion of studies interpreted as equivocal or
4 definitely normal or abnormal. We used patients from two separate sites who underwent a range
5 of stress and imaging protocols. While this increases the generalizability of our results it
6 decreases the precision of our estimate regarding the increase in accuracy for any one
7 combination of camera system and imaging protocol. The MRMC analysis accounts for case
8 variation as well as variations in reader certainty, reader skill, and reader response to AI but more
9 precise evaluation of the impact of explainable AI on physician interpretation could be made in a
10 population imaged with a single camera system and imaging protocol. Additionally, we would be
11 able to make more definitive conclusions about the influence of explainable DL results on reader
12 interpretation if additional readers were involved and all readers interpreted a greater number of
13 studies. As will be the case in clinical practice, interpreters in the study had variable exposure
14 and belief in AI models prior to the study. It is possible that additional experience with utilizing
15 CAD-DL may lead to further improvements in accuracy.

16

17 **CONCLUSION**

18 We demonstrated that overall physician interpretation significantly improved by utilizing the DL
19 predictions compared to the same physicians interpreting without DL. Implementing DL as an
20 aid to physician interpretation significantly improves diagnostic accuracy of myocardial
21 perfusion imaging.

22

1 **ACKNOWLEDGMENTS**

2 We would like to thank all the individuals involved in the collection, processing, and analysis of
3 data in this multicenter registry.

4

5 **SOURCE OF FUNDING**

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

10

11 **DISCLOSURES**

12 Drs. Berman and Slomka and Mr. Kavanagh participate in software royalties for QPS software at
13 Cedars-Sinai Medical Center. Dr. Slomka has received research grant support from Siemens
14 Medical Systems. Drs. Berman, Dorbala, Einstein, and Edward Miller have served as consultants
15 for GE Healthcare. Dr. Einstein has served as a consultant to W. L. Gore & Associates and has
16 served on a speakers bureau for Ionetix. Dr. Dorbala has served as a consultant to Bracco
17 Diagnostics; her institution has received grant support from Astellas. Dr. Di Carli has received
18 research grant support from Spectrum Dynamics and consulting honoraria from Sanofi and GE
19 Healthcare. Dr. Ruddy has received research grant support from GE Healthcare and Advanced
20 Accelerator Applications. Dr. Einstein's institution has received research support from GE
21 Healthcare, Philips Healthcare, Toshiba America Medical Systems, Roche Medical Systems, and

1 W. L. Gore & Associates. Dr. Kuronuma has received funding support from the Society of
2 Nuclear Medicine and Molecular Imaging Wagner-Torizuka Fellowship grant and the Nihon
3 University School of Medicine Alumni Association Research Grant. The remaining authors have
4 no relevant disclosures.

5

6 **KEY POINTS**

7 **QUESTION:**

8 Does an explainable deep learning model, when used as an aid during interpretation, improve
9 physician diagnostic accuracy?

10 **PERTINENT FINDINGS:**

11 In this multiple reader, multiple case interpretation study, access to explainable deep learning
12 results lead to meaningful but variable improvements in the accuracy of physician interpretation
13 of myocardial perfusion imaging. Overall diagnostic performance improved when physicians had
14 access to DL predictions and readers demonstrated improved classification of patients with and
15 without coronary artery disease.

16 **IMPLICATIONS FOR PATIENT CARE:**

17 Explainable deep learning could be implemented as an aid to physician interpretation in order to
18 improve diagnostic accuracy potentially improving patient management and subsequent
19 outcomes.

20

1 REFERENCES

- 2 1. Global Burden of Disease Collaborators. Global, regional, and national age-sex specific mortality
3 for 264 causes of death, 1980-2016. *Lancet*. 2017;390:1151-1210.
- 4 2. Arsanjani R, Xu Y, Hayes SW, et al. Comparison of fully automated computer analysis and visual
5 scoring for detection of coronary artery disease from myocardial perfusion SPECT in a large population. *J*
6 *Nucl Med*. 2013;54:221-228.
- 7 3. Otaki Y, Betancur J, Sharir T, et al. 5-Year Prognostic Value of Quantitative Versus Visual MPI in
8 Subtle Perfusion Defects. *JACC Cardiovasc Imaging*. 2020;13:774-785.
- 9 4. Hachamovitch R, Hayes SW, Friedman JD, Cohen I, Berman DS. Comparison of the short-term
10 survival benefit associated with revascularization compared with medical therapy in patients with no
11 prior coronary artery disease undergoing stress myocardial perfusion single photon emission computed
12 tomography. *Circulation*. 2003;107:2900-2907.
- 13 5. Azadani PN, Miller RJH, Sharir T, et al. Impact of Early Revascularization on Major Adverse
14 Cardiovascular Events in Relation to Automatically Quantified Ischemia. *JACC: Cardiovascular Imaging*.
15 2021;14:644-653.
- 16 6. Betancur J, Commandeur F, Motlagh M, et al. Deep Learning for Prediction of Obstructive
17 Disease From Fast Myocardial Perfusion SPECT. *JACC Cardiovasc Imaging*. 2018;11:1654-1663.
- 18 7. Betancur J, Hu LH, Commandeur F, et al. Deep Learning Analysis of Upright-Supine High-
19 Efficiency SPECT Myocardial Perfusion Imaging for Prediction of Obstructive Coronary Artery Disease: A
20 Multicenter Study. *J Nucl Med*. 2019;60:664-670.
- 21 8. Betancur J, Otaki Y, Motwani M, et al. Prognostic Value of Combined Clinical and Myocardial
22 Perfusion Imaging Data Using Machine Learning. *JACC Cardiovasc Imaging*. 2018;11:1000-1009.
- 23 9. Otaki Y, Singh A, Kavanagh P, et al. Clinical Deployment of Explainable Artificial Intelligence for
24 Diagnosis of Coronary Artery Disease. *JACC Cardiovasc Imaging*. 2021;Epub ahead of print.
- 25 10. Hu LH, Betancur J, Sharir T, et al. Machine learning predicts per-vessel early coronary
26 revascularization after fast myocardial perfusion SPECT. *Eur Heart J Cardiovasc Imaging*. 2020;21:549-
27 559.
- 28 11. Hu LH, Miller RJH, Sharir T, et al. Prognostically safe stress-only single-photon emission
29 computed tomography myocardial perfusion imaging guided by machine learning: report from REFINE
30 SPECT. *Eur Heart J Cardiovasc Imaging*. 2020;22:705-714.
- 31 12. Bradshaw TJ, Boellaard R, Dutta J, et al. Nuclear Medicine and Artificial Intelligence: Best
32 Practices for Algorithm Development. *J Nucl Med*. 2021; Epub ahead of print.
- 33 13. Gerke S, Minssen T, Cohen G. Ethical and legal challenges of artificial intelligence-driven
34 healthcare. *AI Healthcare*. 2020;1:295-336.
- 35 14. Jaremko JL, Azar M, Bromwich R, et al. Canadian Association of Radiologists White Paper on
36 Ethical and Legal Issues Related to Artificial Intelligence in Radiology. *Can Ass Rad J*. 2019;70:107-118.
- 37 15. Pesapane F, Volonté C, Codari M, Sardanelli F. Artificial intelligence as a medical device in
38 radiology: ethical and regulatory issues in Europe and the United States. *Insight Imaging*. 2018;9:745-
39 753.

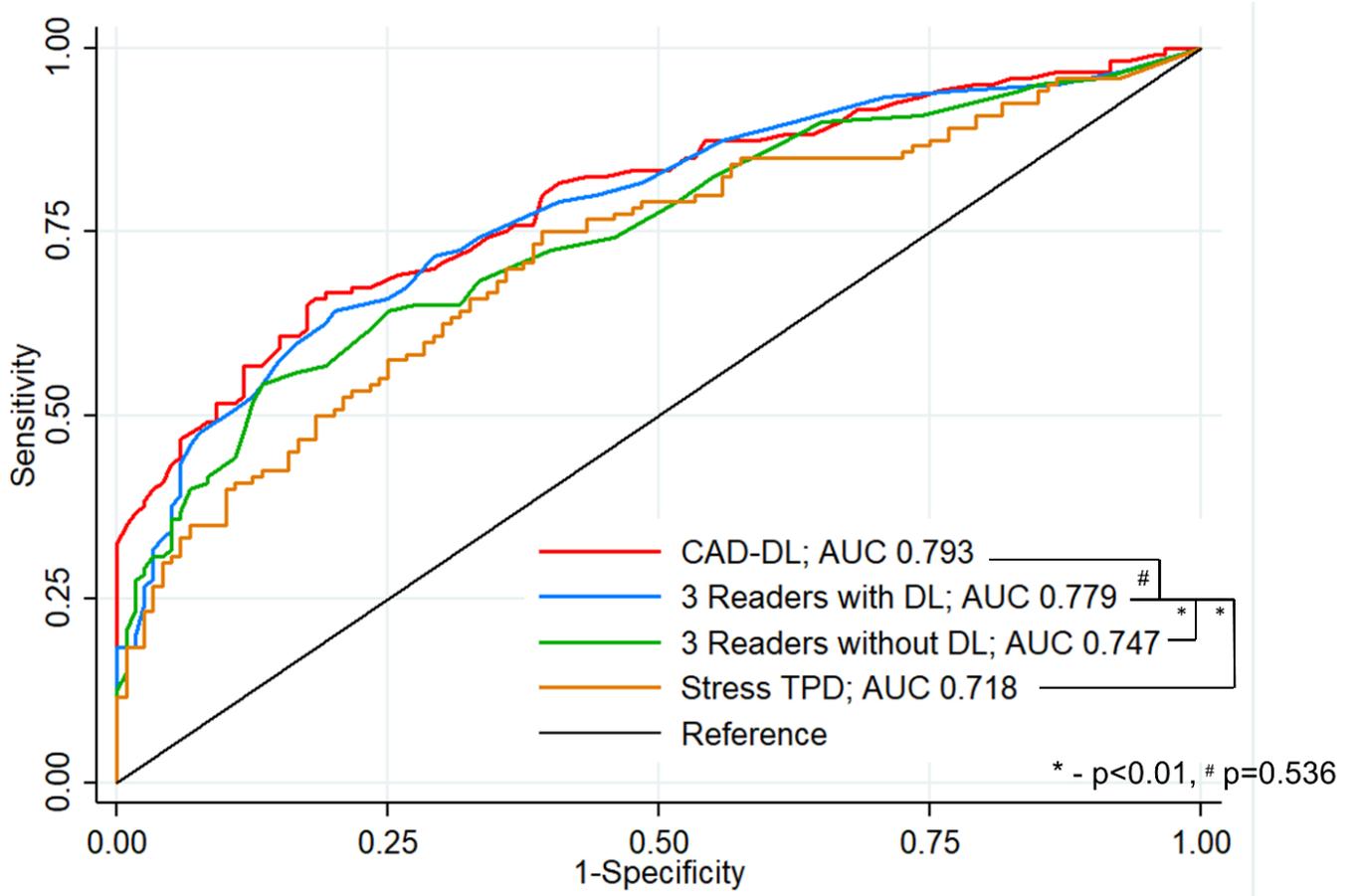
- 1 **16.** Engel TR. Diagnosis of Hypertrophic Cardiomyopathy: Who Is in Charge Here-The Physician or
2 the Computer? *J Am Coll Cardiol.* 2020;75:734-735.
- 3 **17.** Krittanawong C, Johnson KW, Rosenson RS, et al. Deep learning for cardiovascular medicine: a
4 practical primer. *Eur Heart J.* 2019;40:2058-2073.
- 5 **18.** Arsanjani R, Hayes SW, Fish M, et al. Two-position supine/prone myocardial perfusion SPECT
6 (MPS) imaging improves visual inter-observer correlation and agreement. *J Nucl Cardiol.* 2014;21:703-
7 711.
- 8 **19.** Nakazato R, Tamarappoo BK, Kang X, et al. Quantitative upright-supine high-speed SPECT
9 myocardial perfusion imaging for detection of coronary artery disease: correlation with invasive
10 coronary angiography. *J Nucl Med.* 2010;51:1724-1731.
- 11 **20.** Slomka PJ, Nishina H, Berman DS, et al. Automated quantification of myocardial perfusion SPECT
12 using simplified normal limits. *J Nucl Cardiol.* 2005;12:66-77.
- 13 **21.** Karimi-Ashtiani S, Arsanjani R, Fish M, et al. Direct Quantification of Left Ventricular Motion and
14 Thickening Changes Using Rest-Stress Myocardial Perfusion SPECT. *J Nucl Med.* 2012;53:1392-1400.
- 15 **22.** Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations
16 from Deep Networks via Gradient-Based Localization. *IEEE Int Conf Comput Vis.* 27;1:22-29.
- 17 **23.** Hachamovitch R, Berman DS, Kiat H, Cohen I, Friedman JD, Shaw LJ. Value of Stress Myocardial
18 Perfusion Single Photon Emission Computed Tomography in Patients With Normal Resting
19 Electrocardiograms. *Circulation.* 2002;105:823-829.
- 20 **24.** DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated
21 receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988;44:837-845.
- 22 **25.** Gallas BD, Bandos A, Samuelson FW, Wagner RF. A Framework for Random-Effects ROC Analysis:
23 Biases with the Bootstrap and Other Variance Estimators. *Comm Stats.* 2009;38:2586-2603.
- 24 **26.** Clarkson E, Kupinski MA, Barrett HH. A probabilistic model for the MRMC method, part 1:
25 theoretical development. *Acad Radiol.* 2006;13:1410-1421.
- 26 **27.** Arsanjani R, Xu Y, Dey D, et al. Improved accuracy of myocardial perfusion SPECT for the
27 detection of coronary artery disease using a support vector machine algorithm. *J Nucl Med.*
28 2013;54:549-555.
- 29 **28.** Spier N, Nekolla S, Rupprecht C, Mustafa M, Navab N, Baust M. Classification of Polar Maps from
30 Cardiac Perfusion Imaging with Graph-Convolutional Neural Networks. *Sci Rep.* 2019;9:7569.
- 31 **29.** Scott IA. Errors in clinical reasoning: causes and remedial strategies. *BMJ.* 2009;338:b1860.
- 32 **30.** Chiba A, Kudo T, Ideguchi R, et al. Usefulness of an artificial neural network for a beginner to
33 achieve similar interpretations to an expert when examining myocardial perfusion images. *Int J*
34 *Cardiovasc Imaging.* 2021;37:2337-2343.

35

36

1 **FIGURES**

2

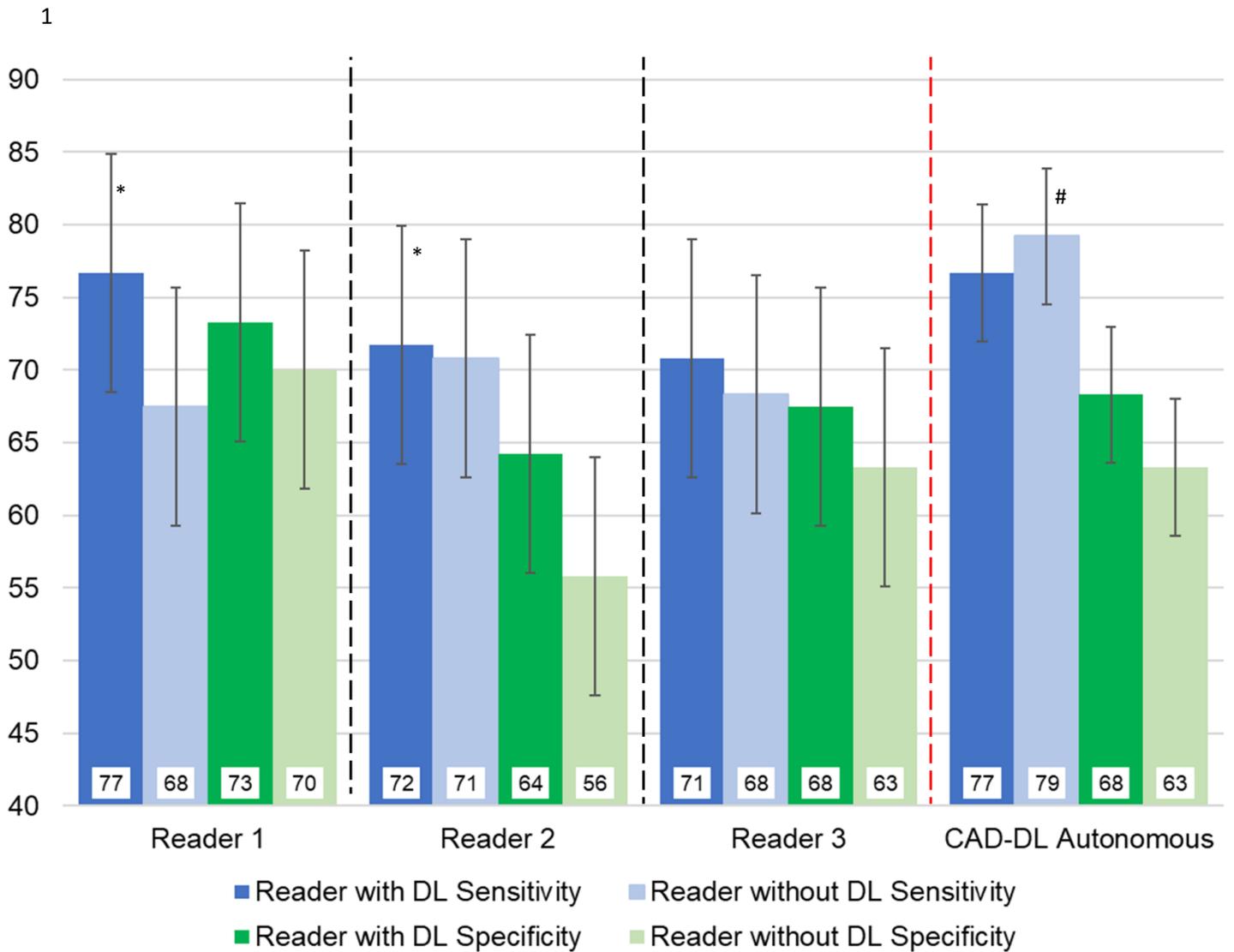


3

4 Figure 1: Diagnostic accuracy of stress perfusion for obstructive coronary artery disease (CAD).
5 Summed stress scores for all readers were averaged to determine reader accuracy with and
6 without deep learning (DL). CAD-DL are results from the DL model when operating
7 autonomously. AUC – Area under receiver operating characteristic curve, TPD – total perfusion
8 deficit.

9

10



2

3 Figure 2: Sensitivity and specificity of reader interpretation with and without deep learning (DL).

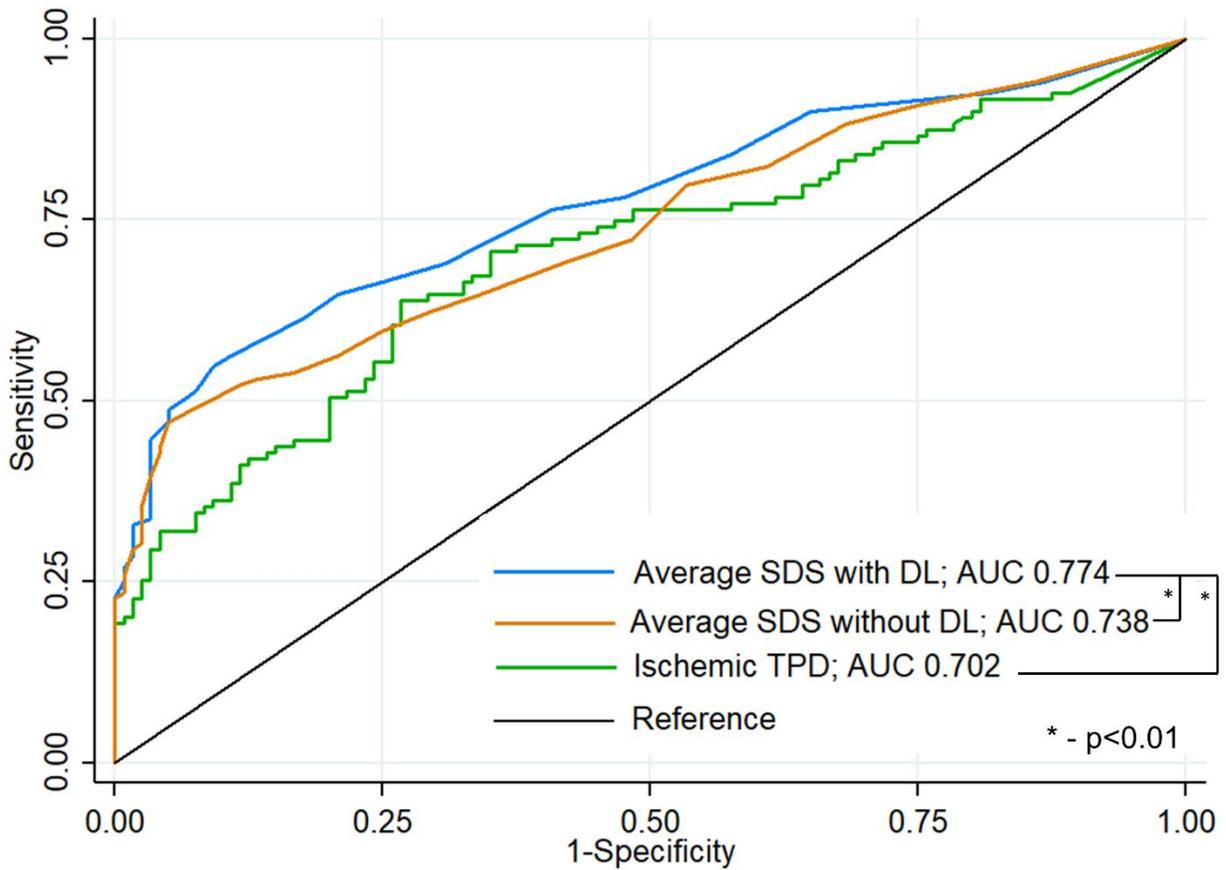
4 Summed stress score >3 was considered abnormal. Two thresholds were established for CAD-

5 DL to match the sensitivity of average reader specificity with (dark bars) and without (light bars)

6 access to DL predictions. * - $p < 0.05$ reader with DL compared to reader without DL, # - $p < 0.001$

7 CAD-DL operating autonomously compared to all readers.

8



1

2 Figure 3: Diagnostic accuracy of ischemia for obstructive coronary artery disease (CAD).

3 Summed difference scores for all readers were averaged to determine reader accuracy with and
 4 without deep learning (DL). AUC – Area under receiver operating characteristic curve, SDS –
 5 summed difference score, TPD – total perfusion deficit.

6

1 **TABLES**

	No obstructive CAD (n=120)	Obstructive CAD (n=120)	p-value
Age (years)	62(55 – 69)	70(62 – 76)	<0.001
Male Sex	65(54.2)	91(75.8)	0.001
BMI (kg/m ²)	28.2(24.8 – 31.8)	27.0(24.0 – 31.3)	0.278
Past Medical History			
Hypertension	78(65.0)	93(77.5)	0.045
Diabetes	36(30.0)	40(33.3)	0.677
Dyslipidemia	54(45.0)	84(70.0)	<0.001
Family History	33(27.5)	33(27.5)	1.000
Smoking	21(17.5)	23(19.2)	0.868
Exercise Stress	53(44.2)	45(37.5)	0.358
Imaging Protocol			0.878
Stress-Rest	77(64.2)	81(67.5)	
Rest-Stress	32(26.7)	29(24.2)	
Stress Only	11(9.2)	10(8.3)	
Left Ventricular ejection fraction (%)	63(54 – 72)	63(53 – 69)	0.356
Stress TPD (%)	3.5(1.8 – 7.2)	8.4(4.3 – 16.3)	<0.001

2 Table 1: Population characteristics stratified by presence of obstructive coronary artery disease
3 (CAD). Categorical values are shown as number (frequency) and continuous values are shown as
4 median (interquartile range). TPD – total perfusion deficit.

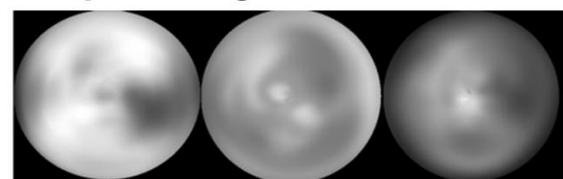
5

5-point scale likelihood without CAD-DL	Reclassified 5-point scale likelihood with CAD-DL					Reclassified likelihood (%)		
	Normal	Probably normal	Equivocal	Probably Abnormal	Definitely Abnormal	Increased likelihood	Decreased likelihood	Net correctly Reclassified
Obstructive CAD (n=360)								
Normal	22	4	2	0	0	47 (13.1%)	25 (6.9%)	6.1% (95% CI 1.4% - 10.3%)
Probably normal	4	34	10	3	3			
Equivocal	0	6	10	14	0			
Probably Abnormal	0	1	4	24	11			
Definitely Abnormal	0	0	0	10	198			
No Obstructive CAD (n=360)								
Normal	64	11	0	0	0	40 (11.1%)	80 (22.2%)	11.1% (95% CI 4.8% - 16.8%)
Probably normal	18	84	11	5	2			
Equivocal	1	13	11	6	1			
Probably Abnormal	0	9	15	29	4			
Definitely Abnormal	1	2	6	15	52			
						NRI +17.2% (95% CI 9.2% – 24.4%, p<0.001)		

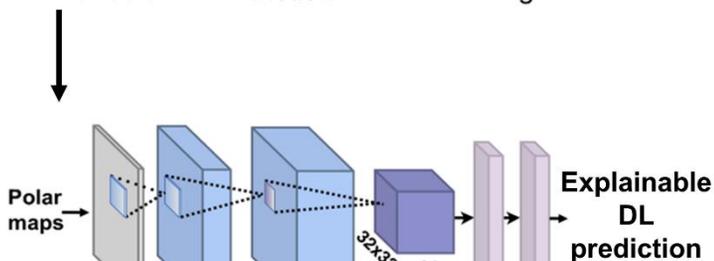
2 Table 2: Net-reclassification of patients with the use of CAD-DL (explainable deep learning model) compared to the same readers
3 without CAD-DL. Blue indicates studies were re-classified in the correct direction and red indicates studies were reclassified in the
4 incorrect direction. CAD – coronary artery disease, CI – confidence interval, NRI – net reclassification index.

1 GRAPHICAL ABSTRACT

Deep learning architecture

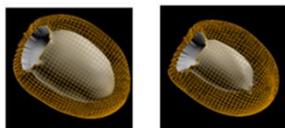


Perfusion Motion Thickening



Input Layer
Convolutional Layers
MaxPool Layer
Fully Connected Layers

Age
Gender



EDV ESV

Explainable
DL
prediction

Reader
without DL



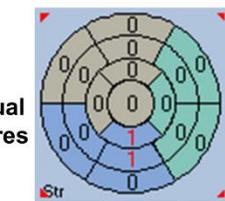
Clinical data



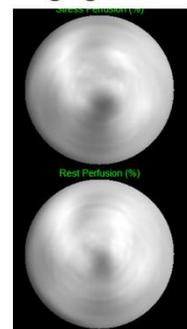
Stress
results



Visual
scores



Imaging results



Reader
With DL



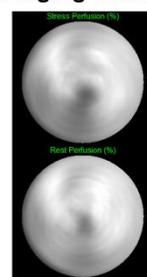
Clinical data



Stress
results

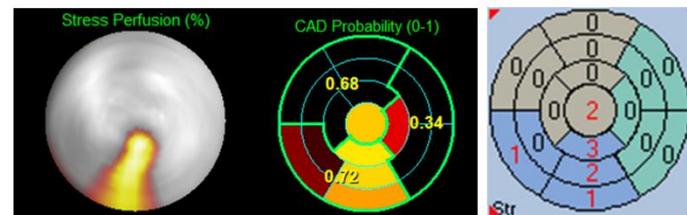


Imaging Results



Explainable DL prediction

Global CAD Risk prediction

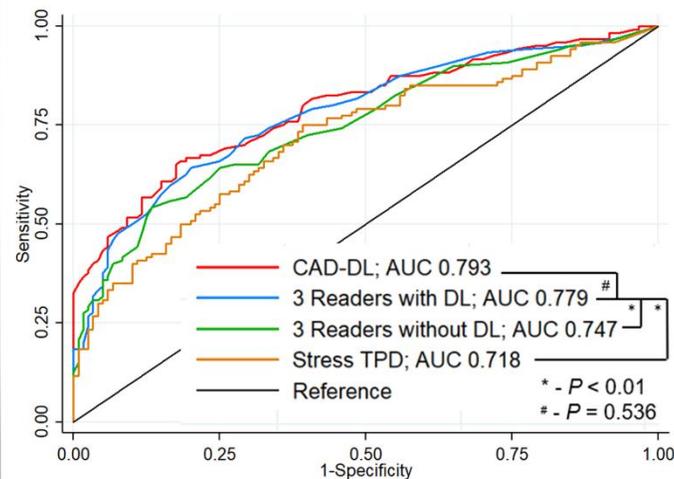


Attention Map

Probability Map

Visual
scores

Diagnostic accuracy for obstructive CAD



1 **SUPPLEMENTAL MATERIAL**

2 **SUPPLEMENTAL METHODS**

3 Invasive Coronary Angiography

4 Invasive coronary angiography was performed according to standard clinical protocols. All
5 coronary angiograms were visually interpreted by an on-site cardiologist. Stenosis $\geq 50\%$ in the
6 left main artery, or $\geq 70\%$ in the left anterior descending artery (LAD), left circumflex artery
7 (LCx), or right coronary artery (RCA), was considered significant and used as the gold standard
8 for obstructive CAD. For per-vessel analyses, left main stenosis was attributed to both LAD and
9 LCx vessels.

10 Automated Image Quantification

11 Quality control for all image data sets was performed by experienced core laboratory
12 technologists without knowledge of the clinical data. Automatically generated myocardial
13 contours by Quantitative Perfusion SPECT (QPS) software (Cedars-Sinai Medical Center, Los
14 Angeles, CA) were evaluated. Stress and rest images in the supine positions were analyzed as
15 previously described(19) using total perfusion deficit (TPD)(20), with normal limits matched to
16 camera type and protocol. Global TPD was used for per-patient analysis, and per-vessel TPD
17 was used for the per-vessel analysis.

18

19 Statistical Analysis

20 Categorical variables are presented as number (frequency) and continuous variables as
21 mean \pm SD or median (interquartile range [IQR]) as appropriate. Categorical variables were

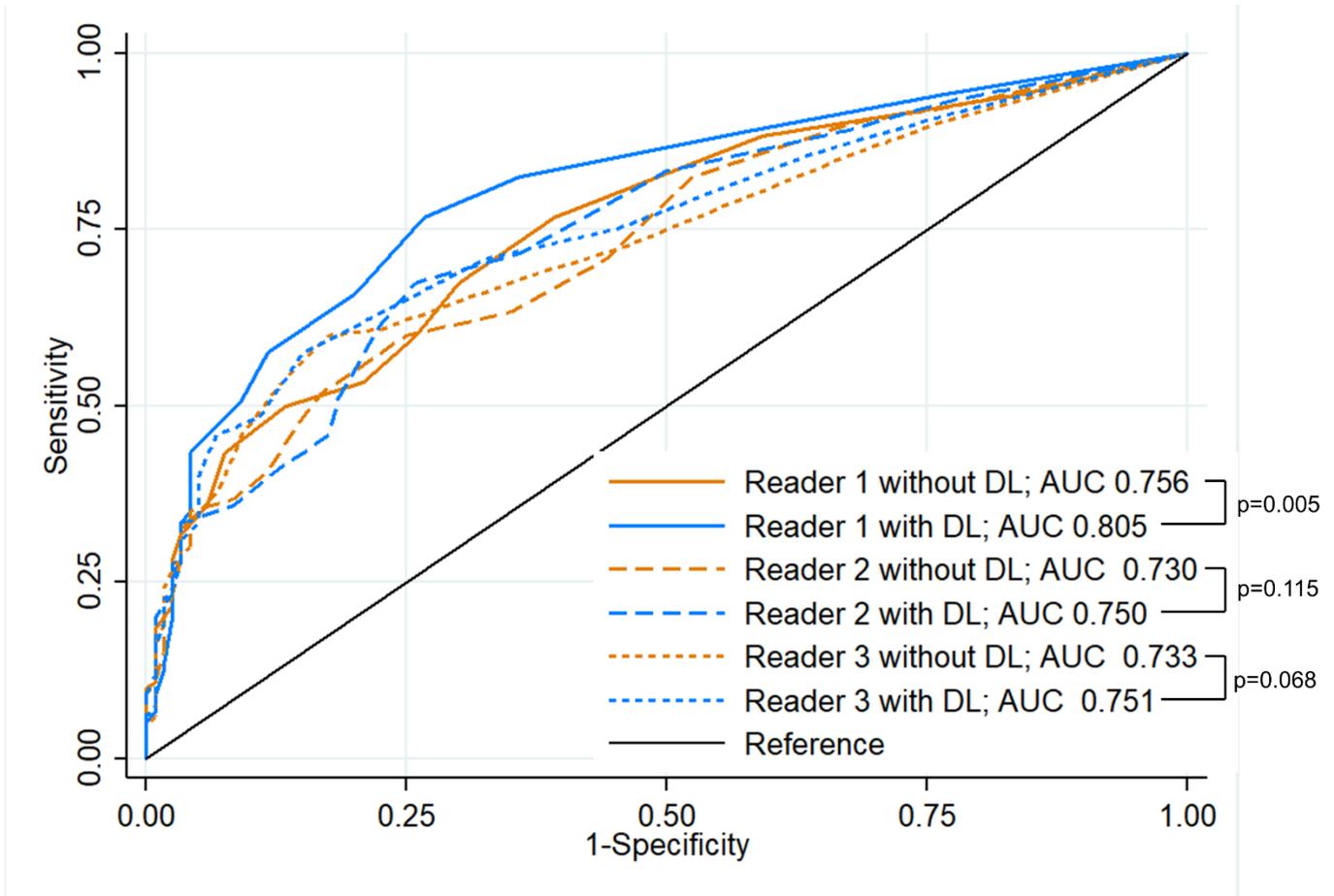
1 compared with chi-square or Fisher exact test as appropriate. Continuous variables were
2 compared with t-tests if normally distributed and with Wilcoxon rank sum test otherwise.

3 In secondary analyses, we also assessed diagnostic accuracy of SDS and overall reader
4 diagnosis with and without DL. Categorical net reclassification improvement (NRI) was calculated
5 based on overall physician diagnosis. All tests were two-sided and a p-value <0.05 was considered
6 significant. Analyses were performed with Stata version 14 (Stata Corp, College Station, TX) and
7 the PredictABEL package (version 1.2-4) and iMRMC package (version 1.2.3) in R (version 4.0.5).

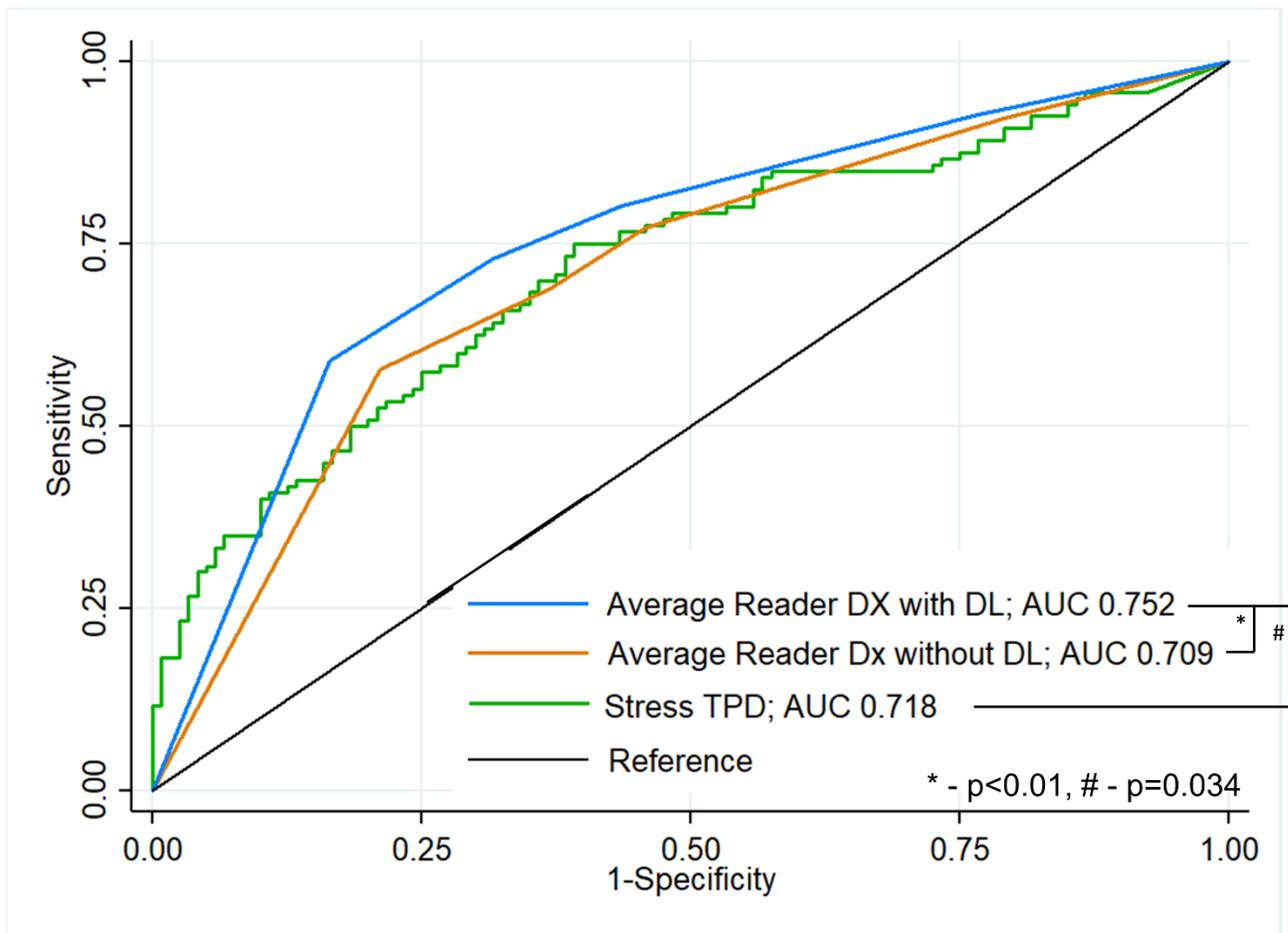
8

1 **SUPPLEMENTAL FIGURES**

2



1



2

3 Supplemental Figure 2: Diagnostic accuracy for obstructive coronary artery disease (CAD).

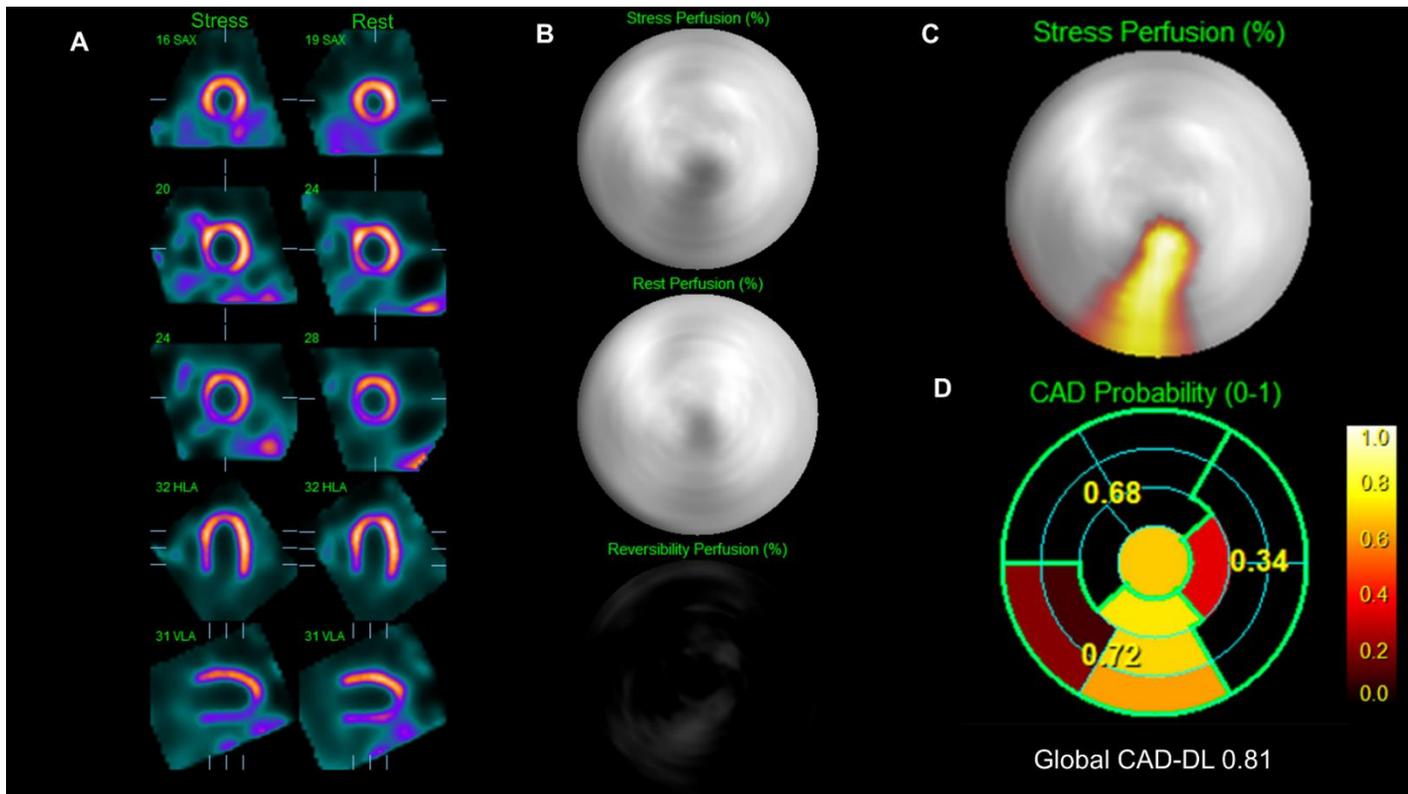
4 Each reader was treated as a separate observation with final diagnosis, using a 5-point scale,
5 compared with and without access to deep-learning (DL) results. AUC – Area under receiver
6 operating characteristic curve, TPD – total perfusion deficit.

7

8

9

1

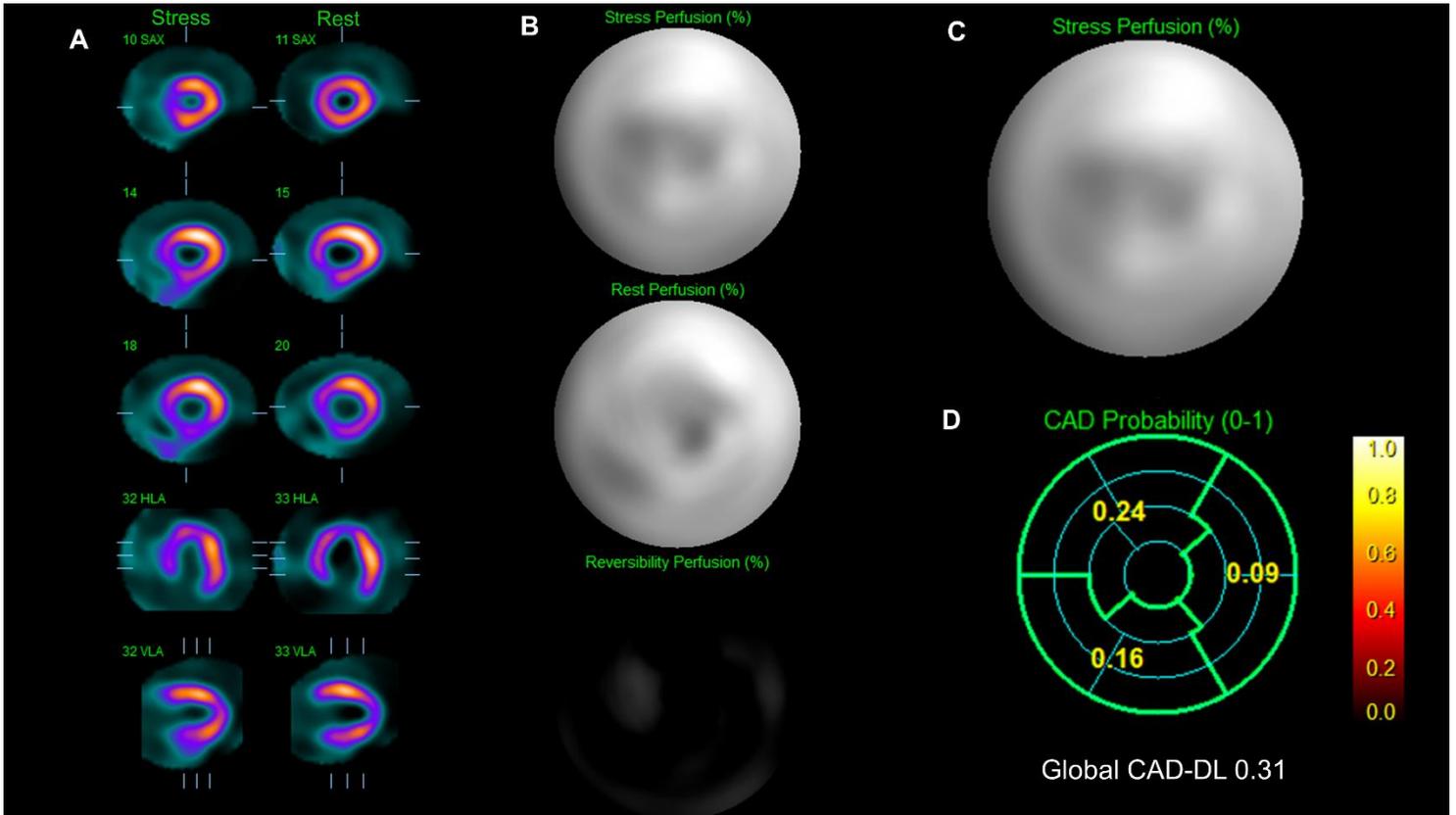


2

3 Supplemental Figure 3: Case example. Example of a case where all 3 physicians increased their
4 segmental scores in a 72-year-old man with obstructive coronary artery disease (CAD). Stress
5 and rest myocardial perfusion are shown in short and long-axis (A) and polar maps (B). The
6 attention map (C) shows that the DL prediction identified an abnormality in the inferior wall,
7 with a correspondingly high likelihood of CAD in the RCA distribution on the CAD probability
8 map (D). The patient had an 80% distal RCA and 70% mid-LAD lesion.

9

1

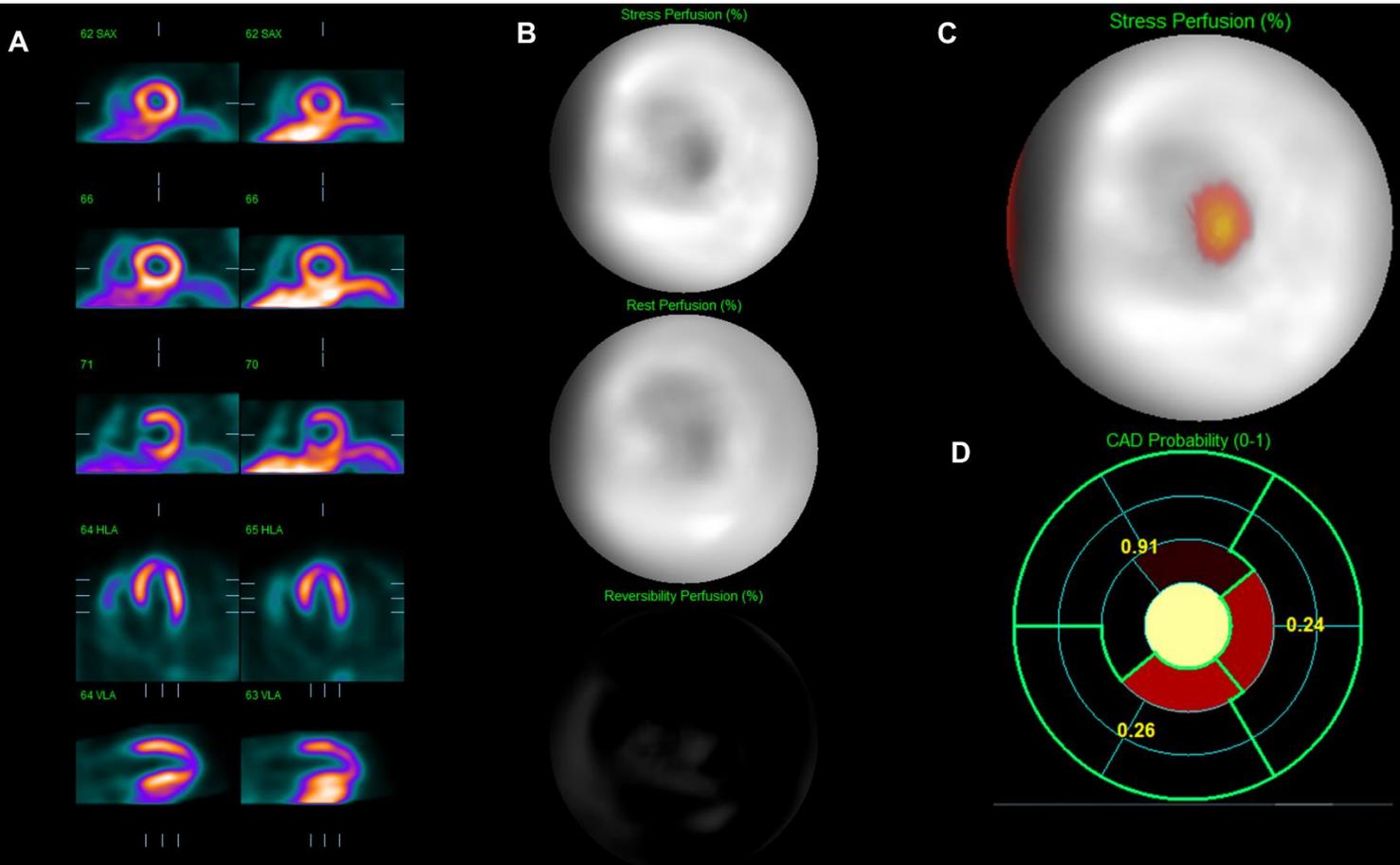


2

3 Supplemental Figure 4: Case Example. Example of a case where all 3 physicians decreased their
4 segmental scores in a 69-year-old woman without obstructive coronary artery disease (CAD).
5 Stress and rest myocardial perfusion are shown in short and long-axis (A) and polar maps (B).
6 The attention map (C) does not highlight any areas because the predicted likelihood of
7 obstructive CAD was low. This corresponds to a low likelihood of obstructive CAD in all
8 segments on the CAD probability map (D). The patient had non-obstructive CAD.

9

1

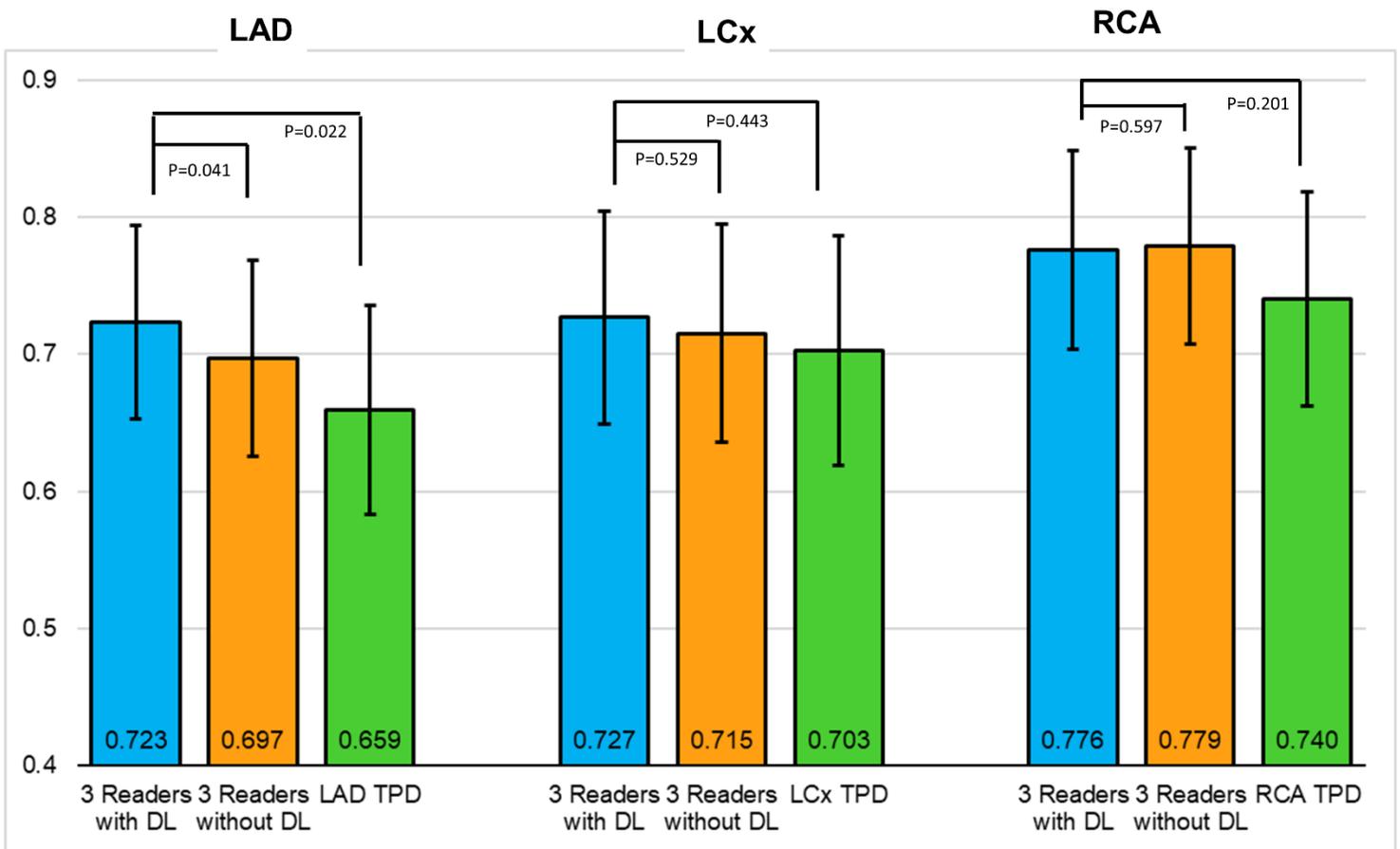


2

3 Supplemental Figure 5: Example of a case where deep-learning (DL) probability was high
4 without consistently high physician summed stress score. Stress and rest myocardial perfusion
5 are shown in short and long-axis (A) and polar maps (B). The attention map (C) shows that the
6 DL prediction identified an abnormality in the apex, with a correspondingly high likelihood of
7 CAD in the LAD distribution on the CAD probability map (D). The patient had an 80% distal-
8 LAD lesion. Physician summed stress scores with DL were 4, 2, and 2.

9

1



2

3 Supplemental Figure 6: Diagnostic accuracy for obstructive coronary artery disease per vessel.

4 DL – deep learning, LAD – left anterior descending, LCx – left circumflex, RCA – right

5 coronary artery, TPD – total perfusion deficit.

6

7

1

	Conventional Camera System (n=80)	Solid-State Camera System (n=160)	p-value
Age (years)	66 (59 – 72)	65 (58 – 73)	0.858
Male Sex	50 (62.5)	106 (66.3)	0.569
BMI (kg/m ²)	29.8 (24.6 – 34.5)	26.7 (24.0 – 30.5)	<0.001
Past Medical History			
Hypertension	59 (73.8)	112 (70.0)	0.650
Diabetes	30 (37.5)	46 (28.8)	0.187
Dyslipidemia	45 (56.3)	93 (58.1)	0.784
Family History	19 (23.8)	47 (29.4)	0.443
Smoking	24 (30.0)	20 (12.5)	0.001
Exercise Stress	25 (31.3)	73 (45.6)	0.037
Imaging Protocol			<0.001
Stress-Rest	25 (31.3)	133 (83.1)	
Rest-Stress	53 (66.3)	8 (5.0)	
Stress Only	2 (2.5)	19 (11.9)	
Left Ventricular Ejection Fraction (%)	63 (54 – 72)	63 (53 – 69)	0.365
Stress TPD (%)	4.2 (1.8 – 9.7)	6.0 (2.8 – 11.5)	0.036
Obstructive CAD	40 (50.0)	80 (50.0)	1.000

2

3 Supplemental Table 1: Population characteristics stratified by camera system. Categorical values

4 are shown as number(frequency) and continuous values are shown as median (interquartile

5 range). CAD – coronary artery disease, TPD – total perfusion deficit.

Patients with scores increased by all readers						
Camera	Age	Sex	Stress TPD	Global CAD-DL score	Highest regional CAD-DL score	CAD Description
Conventional	60	Male	4.9	0.48	0.41 (LAD)	70% LAD
Conventional	59	Male	5.4	0.87	0.61 (RCA)	100% LCx, 50% RCA
CZT	78	Male	6.3	0.74	0.46 (LAD)	80% LAD, 80% LCx, 90% RCA
CZT	72	Male	10.5	0.81	0.72 (RCA)	80% RCA and 70% LAD
CZT	48	Male	4.6	0.63	0.41 (LCx)	100% LAD, 100% LCx, 80% RCA
CZT	69	Female	1.0	0.55	0.44 (RCA)	80% LAD
CZT	78	Male	10.1	0.87	0.64 (LAD)	70% LAD, 90% D2
CZT	68	Male	24.6	0.96	0.82 (LAD)	90% LAD, 90% LCx, 70% RCA
Conventional	63	Male	3.3	0.62	0.56 (LAD)	Minimal luminal irregularities
CZT	47	Male	8.7	0.71	0.58 (RCA)	No CAD
CZT	74	Male	2.3	0.37	0.27 (LAD)	30% LAD
Patients with scores decreased by all readers						
Camera	Age	Sex	Stress TPD	Global CAD-DL score	Highest regional CAD-DL score	CAD Description
Conventional	59	Male	2.0	0.21	0.13 (LAD)	Minimal luminal irregularities
Conventional	56	Female	0.0	0.33	0.21 (LCx)	No CAD
CZT	60	Female	3.7	0.41	0.26 (RCA)	No CAD
CZT	63	Female	2.3	0.28	0.15 (LAD)	No CAD

CZT	73	Male	3.6	0.47	0.29 (RCA)	Minimal luminal irregularities
CZT	59	Male	5.4	0.29	0.17 (LAD)	None
Conventional	69	Male	2.0	0.21	0.13 (LAD)	Minimal luminal irregularities
CZT	67	Female	8.4	0.36	0.21 (RCA)	Minimal luminal irregularities
Conventional	63	Female	3.3	0.36	0.28 (LAD)	No CAD
CZT	69	Female	12.1	0.31	0.24 (LAD)	40% LAD
CZT	34	Female	18.9	0.26	0.18 (RCA)	100% RPL
CZT	64	Female	7.6	0.37	0.21 (RCA)	100% RCA
CZT	74	Female	8.3	0.26	0.16 (LAD)	50% LM, 90% LAD, 90% LCX, 100% RCA

1

2 Supplemental Table 2: Summary of cases where all readers changed scores when interpreting
3 with CAD-DL predictions. CAD – coronary artery disease, CZT – cadmium zinc telluride, D2 –
4 second diagonal, LAD – left anterior descending, LCx – left circumflex, LM – left main, RCA –
5 right coronary artery, RPL – right posterolateral.

6

7

8

LAD	Reclassified Score in Vascular Territory				Reclassified likelihood (%)		
5-point scale likelihood without CAD-DL	0	1 or 2	3 or 4	≥ 5	Increased likelihood	Decreased likelihood	Net correctly Reclassified
Obstructive LAD (n=252)							
0	70	14	1	0	26 (10.3%)	9 (3.6%)	6.7%
1 or 2	1	39	11	0			
3 or 4	0	6	20	10			
≥ 5	0	0	2	78			
No Obstructive LAD (n=468)							
0	237	32	2	1	50 (10.6%)	40 (8.5%)	-2.1%
1 or 2	16	85	10	1			
3 or 4	1	14	29	4			
≥ 5	0	3	6	27	NRI + 4.6%		
LCx	Reclassified Score in Vascular Territory				Reclassified likelihood (%)		
5-point scale likelihood without CAD-DL	0	1 or 2	3 or 4	≥ 5	Increased likelihood	Decreased likelihood	Net correctly Reclassified
Obstructive LCx (n=165)							
0	71	6	0	0	12 (7.3%)	9 (5.5%)	1.8%
1 or 2	4	19	4	0			
3 or 4	0	1	4	2			
≥ 5	0	1	3	50			
No Obstructive LCx (n=555)							
0	419	15	1	0	21 (3.8%)	32 (5.8%)	2.0%
1 or 2	16	57	3	1			
3 or 4	4	10	11	1			
≥ 5	0	1	1	15	NRI + 3.8%		
RCA	Reclassified Score in Vascular Territory				Reclassified likelihood (%)		

5-point scale likelihood without CAD-DL	0	1 or 2	3 or 4	≥ 5	Increased likelihood	Decreased likelihood	Net correctly Reclassified
Obstructive LCx (n=165)							
0	71	6	0	0	12 (7.3%)	9 (5.5%)	1.8%
1 or 2	4	19	4	0			
3 or 4	0	1	4	2			
≥ 5	0	1	3	50			
No Obstructive LCx (n=555)							
0	419	15	1	0	21 (3.8%)	32 (5.8%)	2.0%
1 or 2	16	57	3	1			
3 or 4	4	10	11	1			
≥ 5	0	1	1	15	NRI + 3.8%		

Supplemental Table 3: Net-reclassification by vascular territory. CAD-DL – explainable deep learning model, LAD – left anterior descending, LCx – left circumflex, NRI – net reclassification index, RCA – right coronary artery.