Title: Deep Learning Analysis of Upright-Supine High-Efficiency SPECT Myocardial Perfusion Imaging for Prediction of Obstructive Coronary Artery Disease: A Multicenter Study

Short Title: Deep learning analysis of U/S D-SPECT

Authors: Julian Betancur, PhD [1], Lien-Hsin Hu, MD [1], Frederic Commandeur, PhD [1], Tali Sharir, MD [2,3], Andrew J. Einstein, MD, PhD [4,5], Mathews B. Fish, MD [6], Terrence D. Ruddy, MD [7], Philipp 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], Guido Germano, PhD [1], Yuka Otaki, MD [1], Joanna Liang, BA [1], Balaji K. Tamarappoo, MD [1], Damini Dey, PhD [1], Daniel S. Berman, MD, FACC [1], Piotr J. Slomka, PhD, FACC [1]

Affiliations:

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

[2] Department of Nuclear Cardiology, Assuta Medical Centers, Tel Aviv, Israel

[3] Ben Gurion University of the Negev, Beer Sheba, Israel

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

[5] 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

Corresponding (senior) Author

Piotr Slomka, PhD, FACC

Cedars-Sinai Medical Center

8700 Beverly Boulevard, Ste. A047N

Los Angeles, California 90048

Phone: 310-423-4348

Fax: 310-423-0173

Email: Piotr.Slomka@cshs.org

First Author

Julian Betancur, PhD

Postdoctoral Scientist

Cedars-Sinai Medical Center

8700 Beverly Boulevard

S. Mark Taper Foundation Imaging Center, A238

Los Angeles, California 90048

Phone: 310-423-4201

Fax: 310-423-0173

Email: Julian.Betancur@cshs.org

Total Word Count: 5000

ABSTRACT

Combined analysis of SPECT myocardial perfusion imaging (MPI) performed with a solid-state camera on patients in two positions (semi-upright, supine) is routinely used to mitigate attenuation artifacts. We evaluated the prediction of obstructive disease from combined analysis of semi-upright and supine stress MPI by deep learning (DL) as compared to standard combined total perfusion deficit (TPD). **Methods:** 1160 patients without known coronary artery disease (64% males) were studied. Patients underwent stress ^{99m}Tc-sestamibi MPI with new generation solid-state SPECT scanners in four different centers. All patients had on-site clinical reads and invasive coronary angiography correlations within six months of MPI. Obstructive disease was defined as ≥70% narrowing of the 3 major coronary arteries and ≥50% for the left main coronary artery. Images were quantified at Cedars-Sinai. The left ventricular myocardium was segmented using standard clinical nuclear cardiology software. The contour placement was verified by an experienced technologist. Combined stress TPD was computed using gender- and camera-specific normal limits. DL was trained using polar distributions of normalized radiotracer counts, hypoperfusion defects and hypoperfusion severities and was evaluated for prediction of obstructive disease in a novel leave-one-center-out cross-validation procedure equivalent to external validation. During the validation procedure, four DL models were trained using data from three centers and then evaluated on the one center left aside. Predictions for each center were merged to have an overall estimation of the multicenter performance. **Results:** 718 (62%) patients and 1272 of 3480 (37%) arteries had obstructive disease. The area under the receiver operating characteristics curve for prediction of disease on a per-patient and per-vessel basis by DL was higher than for combined TPD (per-patient: 0.81 vs 0.78, per-vessel: 0.77 vs 0.73, P<0.001). With the DL cutoff set to exhibit the same specificity as the standard cutoff for combined TPD, per-

patient sensitivity improved from 61.8% (TPD) to 65.6% (DL) (P<0.05), and per-vessel sensitivity improved from 54.6% (TPD) to 59.1% (DL) (P<0.01). With threshold matched to specificity of normal clinical read (56.3%) DL had sensitivity 84.8% vs 82.6% for on-site clinical read (P=0.3).

Conclusion: Deep learning improves automatic interpretation of MPI as compared to current quantitative methods.

Keywords: obstructive coronary artery disease, SPECT myocardial perfusion imaging, deep learning, convolutional neural network, total perfusion deficit.

INTRODUCTION

SPECT myocardial perfusion imaging (MPI) is widely used for the diagnosis of coronary artery disease (CAD), with more than 6.3 million scans performed annually in North America (*1*). High-efficiency SPECT scanners equipped with solid-state detectors and specialized collimators have dramatically improved count sensitivity and image resolution (*2*). They enable shorter acquisition times that facilitate imaging patients in multiple positions to assess image artifacts, or alternatively the implementation of low radiation protocols by performing standard time acquisitions or stress-only protocols (*3*).

Machine learning is the capability of artificial intelligence systems to acquire their own knowledge from raw data. Recent research has demonstrated increased capabilities of deep convolutional neural networks (often referred to as deep learning-[DL]) to solve challenging tasks such as classification and image segmentation (*4*). In contrast to conventional machine learning that requires pre-defined statistics to be computed in advance (*5,6*), convolutional networks are connected directly to image pixels and learn image statistics in a self-taught manner therefore processing the images in their natural form (*4*). DL ability to simultaneously analyze normalized radiotracer counts, hypoperfusion defects and hypoperfusion severities SPECT-MPI maps to predict CAD has been recently demonstrated (*7*).

In this work, we aimed to apply DL to upright and supine images acquired for the same patient on a high-efficiency camera. Physicians need to review these two images for the final interpretation. Simple rule-based computational methods are currently used for combined quantification of upright/supine images (*8*) to mitigate various imaging artifacts, such as attenuation artifacts. In this study, utilizing a large international cohort of high-efficiency MPI

with correlating invasive coronary angiography (ICA) we use DL to improve the interpretation of upright/supine images.

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) (9). The registry contains MPI studies and on-site clinical reads of consecutive patients without known CAD who underwent clinically indicated ICA within 180 days of MPI (7). This information was transferred to the core laboratory (Cedars Sinai) for processing. We analyzed stress MPI images from 1160 patients (64% males) from the registry who underwent upright and supine SPECT MPI (performed on DSPECT system) in 4 nuclear cardiology centers in the USA, between 2008 and 2015. The study was approved by the institutional review board of each center and the requirements to obtain informed consent was waived.

Stress Image Acquisition

^{99m}Tc-sestamibi stress imaging was performed using high-efficiency solid-state D-SPECT scanners (Spectrum-Dynamics, Haifa, Israel) for patients positioned semi-upright and supine (10). Patients underwent either symptom-limited Bruce protocol treadmill exercise testing only (522 [45%]) or pharmacologic stress (638 [55%]), with radiotracer injection at peak exercise or during maximal hyperemia, respectively. Upright and supine stress imaging began 15-60 min after stress and followed fast acquisition protocols –acquisition lasted 4-6 min each. Reconstructed images were generated on-site from the list-mode data by vendor-recommended iterative reconstruction

with resolution recovery as optimized on this scanner (*10*). No attenuation, scatter or motion correction was applied.

Invasive Coronary Angiography

ICA was performed within 180 days of the MPI examination according to standard clinical protocols and routine. All coronary angiograms were visually interpreted by an on-site cardiologist, independent from MPI but not formally blinded to MPI. Luminal diameter narrowing of 50% or greater of the left main artery, or of 70% or greater in the left anterior descending (LAD), left circumflex (LCx) or right coronary arteries (RCA), was considered significant and used as the gold standard for obstructive CAD.

Image Processing

Image datasets were transferred and quantified at Cedars-Sinai Medical Center. Left ventricular (LV) myocardial contours were computed using standard Cedars-Sinai Quantitative Perfusion SPECT software version 2015 (*11*). LV contours were verified by a technologist with >15 years of experience in nuclear cardiology who was blinded to angiographic and clinical findings. When needed, the technolgist corrected the gross initial LV localization, the LV mask, and the valve plane position (*6,12*).

Automated Myocardial Perfusion SPECT Quantification

Polar map samples derived from the raw images by the standard algorithm were used to generate raw polar maps showing radiotracer count distributions normalized to the maximal counts. SPECT images were quantified by gender-specific normal limits deriving upright and supine blackout and total perfusion deficit (TPD) maps—four polar maps in total (*8,13*). Blackout polar maps defined areas of hypoperfusion (*11*) as blacked out samples in the raw map. TPD polar

maps described perfusion deficits—individual hypoperfusion severity—as polar map samples on a 0-4 scale (*13*).

Standard Measures

Standard clinical combined TPD (cTPD) measures per-patient and per-vessel for LAD, LCx and RCA territories were computed using the rule of concomitant location of the perfusion defect (*8*) for comparison to DL. Final on-site clinical reads incorporating all clinical information and imaging data (static, gated, stress, rest), were reported on scale 0-4 (0-normal, 1-probably normal, 2-equivocal, 3-abnormal, 4-definitely abnormal) by experts from each site during clinical reporting. Clinical scores 0-1 were considered as normal, and 2-4 were considered as abnormal for sensitivity/specificity calculation.

Deep Learning

The DL procedure illustrated in Fig. 1 extends previous approach detailed in (*7*) by processing upright and supine polar maps simultaneously. Gender information was included to account for the image differences between males and females on non-attenuation corrected MPI. The model estimates the obstructive per-vessel CAD probability, without the use of pre-defined coronary territories or any assumed subdivision of the input polar map. The maximum per-vessel probability of obstructive CAD was retained as the per-patient score.

Leave-one-center-out cross-validation

To externally validate the expected performance of the DL model for each participant center, we divided the studied population per center (four groups) then trained four DL models with data from three centers and evaluated them on the remaining external center in a leave-one-center-out cross-validation (*14*) (Fig. 2). Therefore, the generalizability of the prediction to new MPI data was determined by evaluating unseen patients from a separate center. External validation aims to assess

the accuracy of a model in patients from a different but plausibly related population (*15*). Externally-validated predictions for each center were subsequently merged which reduces the variability of the estimated performance as compared to an external validation in a single external center.**Implementation**

The DL model was implemented using the convolutional architecture for fast feature embedding DL toolkit (Berkeley Artificial Intelligence Research Lab, USA) in Python programming language 2.7.12 (*7,16*). Model training was performed on graphics processing units in a GeForce GTX 1080 Ti card (Nvidia, Santa Clara CA).

Statistical Analysis

DL and cTPD were compared using pairwise comparisons of the area under the receiver operating characteristic curve (AUC) according to DeLong et al. (*17*). Per-vessel curves were computed by merging LAD, LCx and RCA scores by per-vessel cTPD and by the DL scores. McNemar's test was used to assess differences in sensitivity. The per-patient and per-vessel improvements in sensitivity were computed for the DL thresholds matching the specificity when using previously established diagnostic cutoff values of per-patient cTPD = 3%, and per-vessel cTPD = 1% (*8*). Per-patient sensitivity was also evaluated for DL and cTPD thresholds matching the specificity of clinical readers. A two-tailed *P*-value < 0.05 was considered significant. Statistical calculations were performed in R software version 3.4 (*18*).

RESULTS

Baseline characteristics of the studied population (Table 1) were similar to reported single-center studies, with a higher incidence of diabetes (*8,19*). The prevalence of obstructive CAD

(Table 2), age, incidence of diabetes mellitus and hypertension (Supplemental Tables 1 and 2) were the same across centers. Injected radiotracer activity for stress image acquisition are in Table 3.

Left Ventricle Segmentation

LV contours were manually corrected in 143 (12.3%) upright images and 238 (20.5%) supine images (P<0.0001). Upright image modifications included 142 corrections for LV location, 15 corrections for LV mask and 139 valve plane corrections (*12*). Supine image modifications included 238 LV location, 29 LV mask, and 200 valve plane corrections.

Deep Learning

Overall, the leave-one-center-out cross-validation training/validation loop took less than 30 min using graphics processing units, which includes the creation of four DL models. In the testing phase, the prediction of LAD, LCx and RCA disease could be completed in less than 0.5 seconds per patient using CPU computation.

Per-patient. The merged per-patient AUC by DL was significantly higher than the AUC by cTPD (Fig. 3A). When operating with the same specificity as cTPD with a per-patient threshold of 3%, DL significantly improved the sensitivity of per-patient disease prediction. When operating with the same specificity as normal clinical reads, DL had the same sensitivity for disease prediction as on-site clinical readers, and significantly improved the sensitivity compared to cTPD (Fig. 4).

Per-vessel. The merged per-vessel AUC by DL was significanlty higher than the AUC by per-vessel cTPD (Fig. 3B). When operating with the same specificity as cTPD with a per-vessel threshold of 1%, DL significantly improved the sensitivity of per-vessel disease prediction.

Per-center Analysis. DL resulted in significant improvements of per-patient and per-vessel AUCs as compared to cTPD for all externally validated centers except for per-patient CAD prediction in Center 3 where they were similar (Fig. 3 [bottom]).

Performance Per Subpopulation. DL resulted in significant improvements as compared to cTPD for all studied subpopulations except for per-patient AUC for females and for patients undergoing exercise stress test were DL performed the same (Fig. 5).

Case Examples

Fig. 6 illustrates two cases with obstructive CAD correctly predicted by DL. Fig. 6A shows one case with LAD disease correctly predicted by DL but with normal per-vessel and per-patient cTPD. Fig. 6B shows one case with triple-vessel disease correctly identified by DL but with normal TPD at LAD and LCx territories.

DISCUSSION

In this study, we applied DL to automatically combine upright and supine MPI polar maps and to predict obstructive CAD. The performance of DL was compared to clinically established combined perfusion quantification by upright and supine TPD using the rule of the concomitant location of the defect (*8,13*), and to on-site clinical readers. D-SPECT MPI data was collected in a large multicenter registry, to our knowledge the largest to date with ICA correlations, with the

number of samples per vessel disease similar to the number of samples per category in computer vision datasets or new applications in medicine (*7,20,21*).

In a novel leave-one-center-out cross-validation procedure, equivalent to external validation, we observed that DL from upright and supine polar MPI images outperformed cTPD in the prediction of obstructive disease. The observed gains for CAD prediction were greater than those obtained by attenuation-corrected TPD, or by visual analysis of attenuation-corrected SPECT MPI images (*22*). These gains by DL derive from an improved processing of the same information used by cTPD with no additional testing, radiation, or cost. It should be also kept in mind that there is an upper limit in diagnostic MPI performance measured versus ICA due to physiological constraint of prediction of stenosis from perfusion defect.

We demonstrate that DL matched the sensitivity/specificity of disease prediction by clinical expert reading, even though on-site reader experts made their diagnosis with all imaging data (including rest scans and gated scans) as well as patient's clinical information, all unknown by DL and cTPD. Moreover, experts reported in their own laboratory, while DL was trained on data excluding patients from the tested laboratory. This validation approach conservatively reflects the performance estimate of the generalization of the DL approach to unseen centers for potential external deployment of such technology.

The proposed DL procedure performs an automatic end-to-end prediction of CAD (from the image pixels to the predicted CAD score) of upright and supine MPI. This demonstrates DL flexibility and extends previous single-view approach (*7*). The DL procedure learned image statistics from supine and upright maps and integrated them with gender information to compute a score for obstructive CAD outperforming cTPD prediction. The DL procedure was able to

capture complex relationships that were not easily captured with the rule for upright and supine concomitancy as observed on the clinical examples in Fig. 6.

We observed AUC improvements for males, patients undergoing pharmacological stress, same-day rest/stress, two-day stress/rest or stress-only MPI. We also observed per-patient and per-vessel AUC improvements by DL as compared to cTPD, for each participant center. Some inter-center AUC variability was observed, likely due to the differences in patients between centers. For instance, we observed a higher proportion of patients undergoing the exercise stress test in Center 1 than the other two centers (Supplemental Table 1). The variability may also be due to the differences in imaging protocols for obtaining the upright and supine images among the centers.

The proposed DL method can be easily and immediately deployed clinically. The expected increased performance should be reproduced in new upright/supine data from unseen centers as observed here after rigorous external validation. Potentially, the method can be extended for the simultaneous analysis of prone and supine images, or attenuation-corrected and non-attenuation-corrected images. The proposed method is standalone and utilizes only image data, combining optimally the upright and supine views, which is difficult visually. In comparison to machine learning models combining image and clinical information (*23*), the presented technique does not require entering clinical information by the physician, which is potentially dependent on the automatic access to the electronic health records. The predicted per-vessel probabilities of obstructive CAD could be integrated with the quantitative software, in the form of polar maps (Fig. 6) to aid the clinician in final reporting. The execution time of such DL models (<1 s), allows for routine application of what can be the first practical deployment of DL technology to nuclear medicine and cardiac imaging and one of the first examples for DL application in clinical practice.

This study has several limitations. First, visual stenosis on ICA was used as the gold standard, which is known to overestimate the prevalence of functionally significant disease when compared with fractional flow reserve. Fractional flow reserve measurements were not available in this population, as these are not commonly performed clinically. The accuracy of stenosis interpretation may also differ between centers. The established endpoint interval of 180 days between SPECT MPI and clinical indicated ICA may be considered long. We did employ a 70% diameter stenosis cutoff for the three major coronary arteries to indicate functionally significant lesions (*24*). Second, in this study, we used upright and supine polar maps acquired in D-SPECT cameras as this protocol is specific to this camera. Motion correction was not considered for these scanners as it was not available at any of the participant centers at the time when the images were collected and interpreted; however, the proposed simultaneous processing of two-position images would potentially overcome motion artifacts in one-position images. Third, rest scans and ischemia were not considered; however, in this work, we studied patients without known CAD where traditionally stress variables are used for detection of obstructive disease. It is possible that externally validated machine learning models, also incorporating clinical information into the scoring of obstructive disease (*5*), provide further enhancements in AUC performance, especially for the prediction of prognostic outcomes.

CONCLUSION

DL improves automatic prediction of obstructive coronary artery disease from MPI, as compared to the current standard quantitative method.

DISCLOSURE

ACKNOWLEDGMENTS

REFERENCES

1.   Einstein AJ. Effects of radiation exposure from cardiac imaging: How good are the data? *J Am Coll Cardiol.* 2012;59:553-565.

2.   Slomka PJ, Patton JA, Berman DS, Germano G. Advances in technical aspects of myocardial perfusion SPECT imaging. *J Nucl Cardiol.* 2009;16:255-276.

3.   Slomka PJ, Dey D, Duvall WL, Henzlova MJ, Berman DS, Germano G. Advances in Nuclear Cardiac Instrumentation with a View Towards Reduced Radiation Exposure. *Curr Cardiol Rep.* 2012;14:208-216.

4.   LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521:436-444.

5.   Arsanjani R, Xu Y, Dey D, et al. Improved accuracy of myocardial perfusion SPECT for detection of coronary artery disease by machine learning in a large population. *J Nucl Cardiol.* 2013;20:553-562.

6.   Betancur J, Rubeaux M, Fuchs T, et al. Automatic valve plane localization in myocardial perfusion SPECT/CT by machine learning: Anatomical and clinical validation. *J Nucl Med.* 2017;58:961-967.

7.   Betancur J, Commandeur F, Motlagh M, et al. Deep learning for prediction of obstructive disease from fast myocardial perfusion SPECT: A multicenter study. *JACC Cardiovasc Imaging.* 2018 (online).

8.   Nakazato R, Tamarappoo BK, Kang X, et al. Quantitative upright–supine high-speed SPECT myocardial perfusion imaging for detection of coronary artery disease: Correlation with invasive coronary angiography. *J Nucl Med.* 2010;51:1724-1731.

9.   Slomka PJ, Betancur J, Liang JX, et al. Rationale and design of the registry of fast myocardial perfusion imaging with next generation SPECT (REFINE SPECT). *J Nucl Cardiol.* 2018 (online).

10.   Gambhir SS, Berman DS, Ziffer J, et al. A novel high-sensitivity rapid-acquisition single-photon cardiac imaging camera. *J Nucl Med.* 2009;50:635-643.

11.   Germano G, Kavanagh PB, Slomka PJ, Van Kriekinge SD, Pollard G, Berman DS. Quantitation in gated perfusion SPECT imaging: The Cedars-Sinai approach. *J Nucl Cardiol.* 2007;14:433-454.

12.   Xu Y, Kavanagh P, Fish M, et al. Automated quality control for segmentation of myocardial perfusion SPECT. *J Nucl Med.* 2009;50:1418-1426.

13.   Slomka PJ, Nishina H, Berman DS, et al. Automated quantification of myocardial perfusion SPECT using simplified normal limits. *J Nucl Cardiol.* 2005;12:66-77.

14.   Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol.* 2016;69:245-247.

15.   Bleeker SE, Moll HA, Steyerberg EW, et al. External validation is necessary in prediction research: A clinical example. *J Clin Epidemiol.* 2003;56:826-832.

16.   Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional architecture for fast feature embedding. Paper presented at: MM' 14, Proceedings of the 22nd ACM international conference on Multimedia; November 3-7, 2014; Orlando, FL.

17.   DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics.* 1988;44:837-845.

18.   Team RC. R: A language and environment for statistical computing. *R Foundation for Statistical Computing.* 2017.

19.   Nishina H, Slomka PJ, Abidov A, et al. Combined supine and prone quantitative myocardial perfusion SPECT: Method development and clinical validation in patients with no known coronary artery disease. *J Nucl Med.* 2006;47:51-58.

20.   Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. Paper presented at: 2009 IEEE Conference on Computer Vision and Pattern Recognition; June 2009, 2009.

21.   Yousefi S, Amrollahi F, Amgad M, et al. Predicting clinical outcomes from large scale cancer genomic profiles with deep survival models. *Sci Rep.* 2017;7:11707.

22.   Arsanjani R, Xu Y, Hayes SW, 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;54:221-228.

23.   Betancur J, Otaki Y, Motwani M, et al. Prognostic value of combined clinical and myocardial perfusion imaging data using machine learning. *JACC Cardiovasc Imaging.* 2017 (Epub ahead of print).

24.   Tonino PAL, Fearon WF, De Bruyne B, et al. Angiographic versus functional severity of coronary artery stenoses in the FAME study: Fractional flow reserve versus angiography in multivessel evaluation. *J Am Coll Cardiol.* 2010;55:2816-2821.
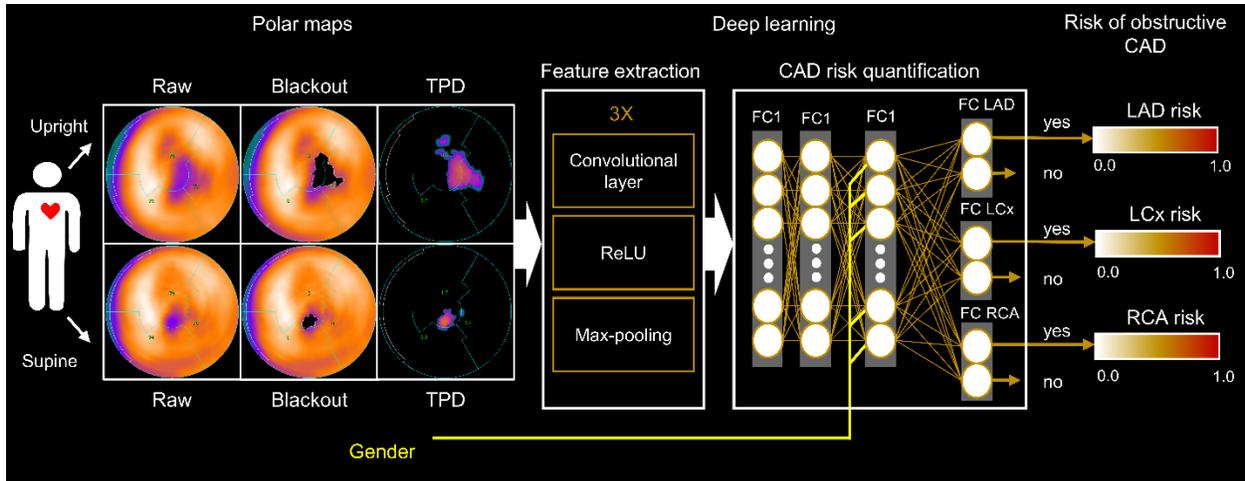
FIGURE 1. Deep learning prediction of obstructive CAD from upright and supine MPI. A deep convolutional neural network trained from obstructive stenosis correlations by invasive coronary angiography was used to simultaneously estimate the probability of obstructive CAD for LAD, LCx and RCA territories from upright and supine polar MPI maps. The maximum probability was retained as the probability of patient disease. CAD: coronary artery disease, FC: fully connected layer, LAD: left anterior descending, LCx: left circumflex, Max-pooling: function that return the maximum value for an image patch, MPI: SPECT myocardial perfusion imaging, RCA: right coronary arteries, ReLU: rectified linear unit, TPD: total perfusion deficit.
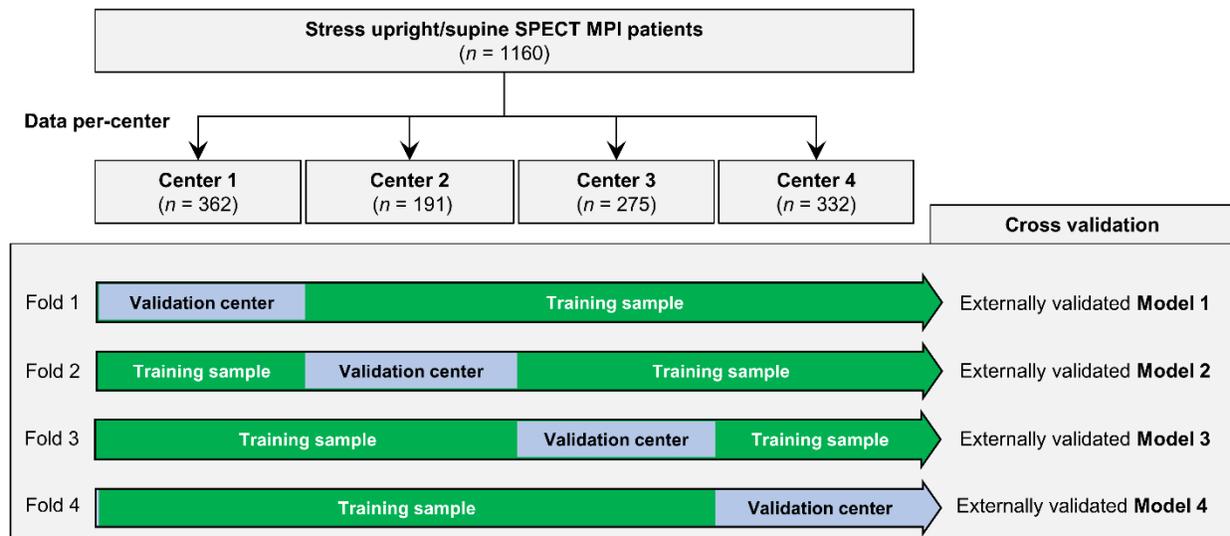
FIGURE 2. Leave-one-center-out cross-validation. Input stress MPI datasets are divided by center (four in total). Four folds are built, each containing a training sample made up of the images from three centers and a validation sample with the images from the remaining center. This procedure allows to externally validate four deep learning models trained separately in each fold.
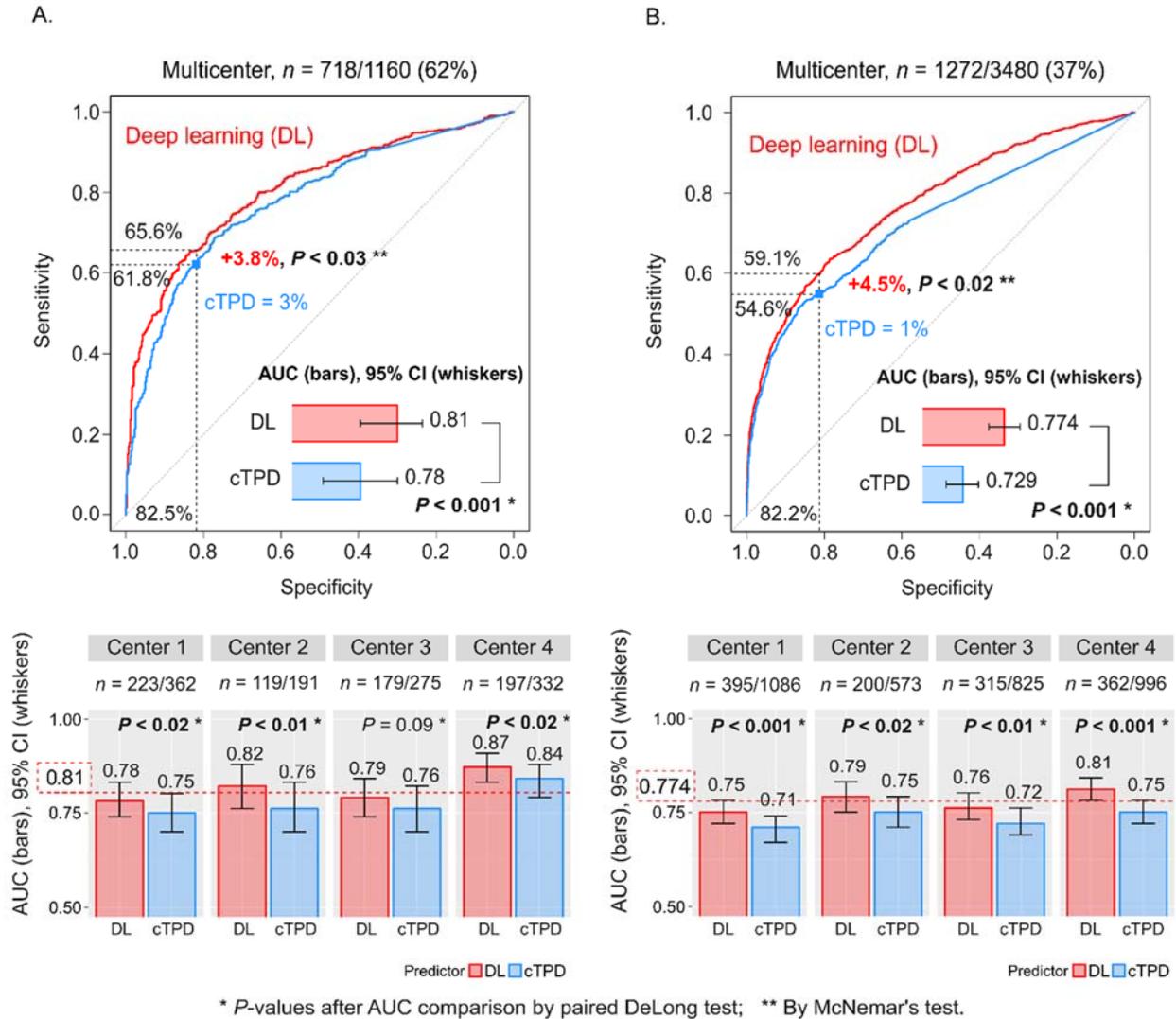
FIGURE 3. DL prediction of obstructive CAD from four externally validated models--with merged data for four centers. Per-patient (A) and per-vessel (B) DL predictions of obstructive CAD from upright and supine images (DL, red) are compared with the prediction of obstructive CAD by combined upright-supine TPD (cTPD, blue). AUC per center was externally validated by using CAD scores from four different DL models (one per center) with each model trained with data from the other three centers. Red dotted line (bottom) shows overall multicenter AUC. AUC: area under ROC curve, DL: deep learning, ROC: receiver operating characteristic, TPD: total perfusion deficit.
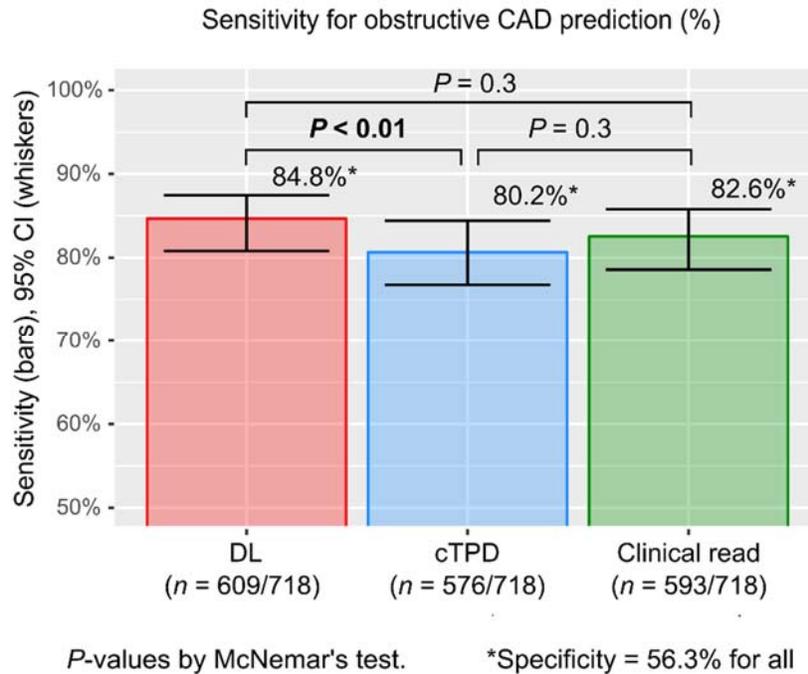
FIGURE 4. Sensitivity for prediction of obstructive CAD. Per-patient DL prediction of obstructive CAD by DL computed from upright and supine MPI (DL, red) had higher sensitivity than the prediction by combined TPD (cTPD, blue), and the same sensitivity as on-site clinical readers (Clinical read, green). DL cutoff was set to 0.29, and cTPD cutoff was set to 0.62% to exhibit the same specificity as normal or probably normal clinical read Acronyms as in Fig. 3.

A.

Predictor ■ DL ■ cTPD

	Female	Male		Non-obese	Obese	
	n = 205/415	n = 513/745		n = 415/645	n = 303/515	
	P = 0.31	**P < 0.001**	Sex	**P < 0.001**	**P < 0.05**	BMI at 30 kg/mt²
	0.77 / 0.76	0.81 / 0.78		0.82 / 0.78	0.8 / 0.78	
	Low	Standard		Exercise	Pharma	
	n = 206/309	n = 512/851		n = 326/522	n = 392/638	
	P < 0.04	**P < 0.01**	Stress dose	P = 0.68	**P < 0.001**	Stress test type
	0.78 / 0.75	0.81 / 0.79		0.84 / 0.83	0.78 / 0.74	

AUC (bars), 95% CI (whiskers)

P-values after AUC comparison by paired DeLong test

B.

Predictor ■ DL ■ cTPD

	Female	Male		Non-obese	Obese	
	n = 341/1245	n = 931/2235		n = 759/1935	n = 513/1545	
	P < 0.001	**P < 0.001**	Sex	**P < 0.001**	**P < 0.001**	BMI at 30 kg/mt²
	0.774 / 0.729	0.757 / 0.721		0.78 / 0.74	0.77 / 0.72	
	Low	Standard		Exercise	Pharma	
	n = 365/927	n = 907/2553		n = 561/1566	n = 711/1914	
	P < 0.01	**P < 0.001**	Stress dose	**P < 0.001**	**P < 0.001**	Stress test type
	0.76 / 0.72	0.78 / 0.73		0.76 / 0.72	0.78 / 0.73	

AUC (bars), 95% CI (whiskers)

P-values after AUC comparison by paired DeLong test

22

FIGURE 5. Prediction of obstructive CAD per-subpopulation. REFINE-SPECT subpopulations were defined by sex (F, M), obesity (Non-obese: BMI < 30 kg/m^2, Obese: BMI ≥ 30 kg/m^2), stress imaging activity (Low: patients undergoing stress-first/stress-only MPI, Standard: patients undergoing rest-first/two-day MPI), and stress test type (Exercise, Pharmacological). The red dotted line shows overall multicenter AUC as reported in Fig. 3. cTPD: combined upright-supine TPD, BMI: body mass index. Other acronyms as in Fig. 3.
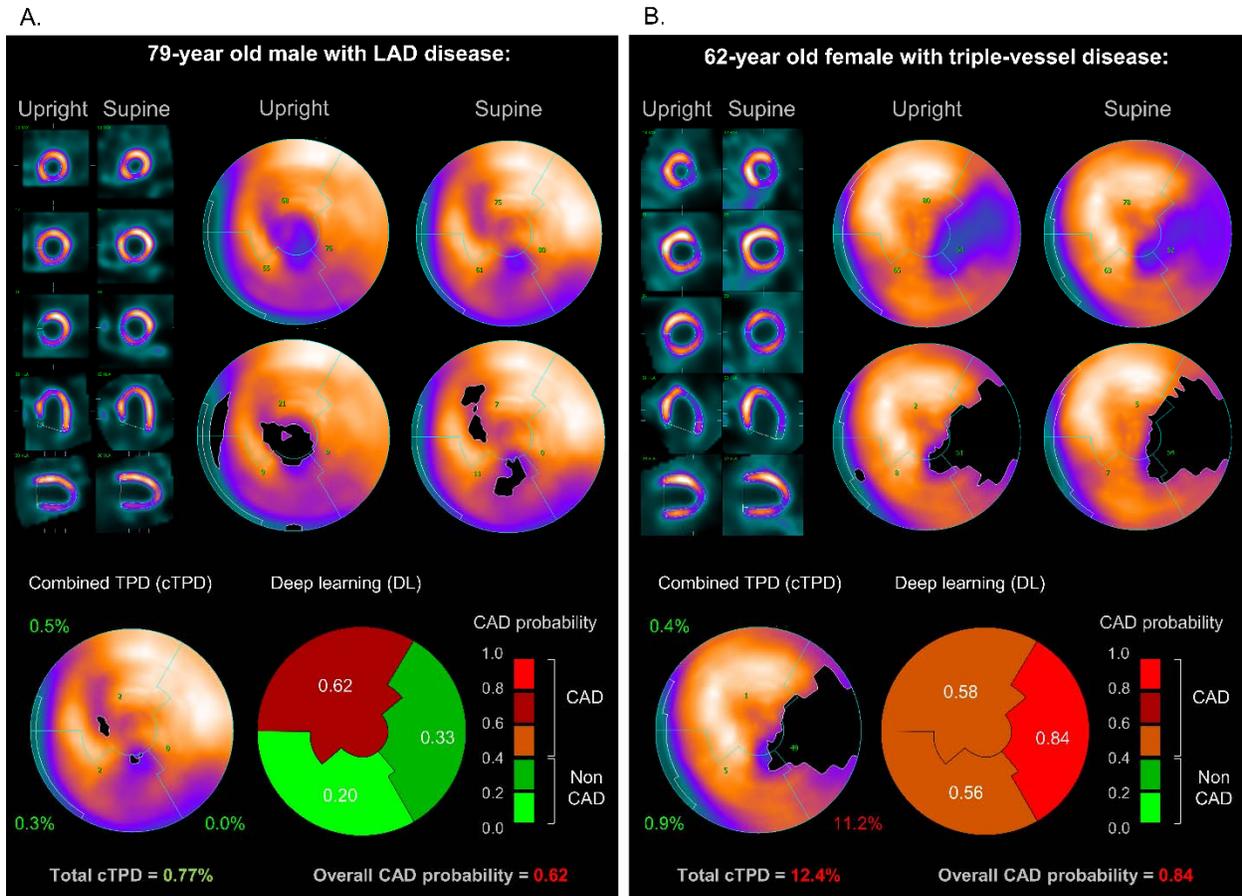
FIGURE 6. Prediction of obstructive CAD from upright and supine stress MPI. The short/long axis views, polar maps depicting normalized radiotracer count distribution and perfusion defects (top), and predictions by cTPD and DL (bottom) are shown for two patients with obstructive CAD. **(A)** In a 79-year-old male (85% proximal LAD stenosis) quantified with normal cTPD (per-patient cTPD < 3% and per-vessel cTPD < 1%), DL correctly identified the LAD disease. The patient had BMI = 30 kg/m^2, diabetes, and underwent exercise stress MPI. **(B)** In a 62-year-old female (70% mid LAD stenosis, 95% proximal LCX stenosis and 80% proximal RCA stenosis) with cTPD abnormal for one vessel only, DL correctly identified the triple-vessel disease. The patient had BMI = 25 kg/m^2, dyslipidemia, family history of cardiac disease and underwent exercise stress MPI. BMI: body mass index. Other acronyms as in Fig. 1.

24

TABLES

Table 1: Baseline characteristics of the studied population.

Characteristic	Overall n=1160	Non-obstructive CAD n=442 (38.1%)	Obstructive CAD n=718 (61.9%)	P-value
Age (years)	64.3±11.5	62.2±12	65.6±11.1	**<0.0001**
Sex				
Male, n (%)	745 (64.2)	232 (52.5)	513 (71.5)	**<0.0001**
Female, n (%)	415 (35.8)	210 (47.5)	205 (28.5)	**<0.0001**
Weight (kg)	88.5±21.9	89.4±23.6	87.9±20.8	=0.278
BMI (kg/m2)	30±6.5	30.8±7.5	29.5±5.7	**<0.01**
Diabetes Mellitus, n (%)	351 (30.3)	121 (27.4)	230 (32)	=0.09
Hypertension, n (%)	841 (72.5)	304 (68.8)	537 (74.8)	**<0.05**
Dyslipidemia, n (%)	780 (67.2)	281 (63.6)	499 (69.5)	**<0.05**
Smoking, n (%)	221 (19.1)	83 (18.8)	138 (19.2)	=0.85
Stress test type				
Exercise, n (%)	522 (45)	196 (44.3)	326 (45.4)	=0.73
Exercise + Pharmacological, n (%)	164 (14.1)	48 (10.9)	116 (16.2)	**<0.05**
Pharmacological, n (%)	474 (40.9)	198 (44.8)	276 (38.4)	**<0.05**
Imaging Protocol				
Stress only, n (%)	48 (4.1)	19 (4.3)	29 (4)	=0.83
Same day stress and rest, n (%)	1073 (92.5)	403 (91.2)	670 (93.3)	=0.18
- Stress-first, n (%)	261 (22.5)	84 (19.0)	177 (24.7)	**<0.05**
- Rest-first, n (%)	812 (70.0)	319 (72.2)	493 (68.7)	**<0.05**
Two-day stress and rest, n (%)	39 (3.4)	20 (4.5)	19 (2.7)	=0.09

CAD: coronary artery disease, BMI: body mass index.

Table 2. Prevalence of obstructive CAD.

Obstructive disease	Prevalence					
	Overall multicenter n=1160	Center 1 n=362	Center 2 n=191	Center 3 n=275	Center 4 n=332	P-value
No disease, n (%)	442 (38.1)	139 (38.4)	72 (37.8)	96 (34.9)	135 (40.7)	=0.54
One-vessel disease, n (%)	321 (27.7)	100 (27.6)	62 (32.5)	79 (28.7)	80 (24.1)	=0.22
Double-vessel disease, n (%)	240 (20.7)	74 (20.4)	33 (17.3)	64 (23.3)	69 (20.8)	=0.48
Triple-vessel disease, n (%)	157 (13.5)	49 (13.6)	24 (11.0)	36 (13.1)	48 (14.5)	=0.93
Per-patient, n (%)	**718 (61.9)**	**223 (61.6)**	**119 (62.3)**	**179 (65.1)**	**197 (59.3)**	=0.54
LAD, n (%)	509 (43.9)	163 (45.0)	84 (44.0)	124 (45.1)	138 (41.6)	=0.78
LCx, n (%)	384 (33.1)	121 (33.4)	60 (31.4)	92 (33.5)	111 (33.4)	=0.96
RCA, n (%)	379 (32.7)	111 (30.7)	56 (29.3)	99 (36)	113 (34.0)	=0.35
Per-vessel (LAD+LCx+RCA), n (%)	**1272/3480 (36.6)**	**395/1086 (36.4)**	**200/573 (34.9)**	**315/825 (38.2)**	**362/996 (36.4)**	=0.65

CAD: coronary artery disease, LAD: left anterior descending, LCx: left circumflex, RCA: right coronary arteries.

Table 3. Radiotracer activity for stress image acquisition.

Image protocol	Injected activity (MBq)
Same-day protocols, n=1073 (92.5%)	
Stress-first protocol, n=261 (24.3%)	213.1 ± 87.3
Rest-first protocol, n=812 (75.7%)	103 ± 384
Two-day protocol, n=39 (3.4%)	682.3 ± 481.44
Stress-only protocol, n=48 (4.1%)	260.2 ± 486.5
Overall, n=1160	**804 ± 494**