

Comparing various AI approaches to traditional quantitative assessment of the myocardial perfusion in [⁸²Rb] PET for MACE prediction

Sacha Bors^{1,3}, Daniel Abler^{2,3}, Matthieu Dietz⁴, Vincent Andrearczyk^{1,3}, Julien Fageot⁵, Marie Nicod-Lalonde^{1,6}, Niklaus Schaefer^{1,6}, Robert DeKemp⁷, Christel H. Kamani⁸, John O. Prior^{1,6,*}, and Adrien Depeursinge^{1,3}

¹Nuclear Medicine and Molecular Imaging Department, Lausanne University Hospital, Lausanne, Switzerland

²Department of Oncology, Lausanne University Hospital, Lausanne, Switzerland

³Institute of Informatics, School of Management, HES-SO Valais-Wallis University of Applied Sciences and Arts Western Switzerland, Sierre, Switzerland

⁴INSERM U1060, CarMeN laboratory, University of Lyon, Lyon, France

⁵AudioVisual Communications Laboratory (LCAV), EPFL, Lausanne, Switzerland

⁶University of Lausanne, Lausanne, Switzerland

⁷University of Ottawa Heart Institute, Division of Cardiology, Ottawa, Ontario, Canada

⁸Department of Cardiology, Lausanne University Hospital, Lausanne, Switzerland

*John.Prior@chuv.ch

ABSTRACT

Assessing the individual risk of Major Adverse Cardiac Events (MACE) is of major importance as cardiovascular diseases remain the leading cause of death worldwide. Quantitative Myocardial Perfusion Imaging (MPI) parameters such as stress Myocardial Blood Flow (sMBF) or Myocardial Flow Reserve (MFR) constitutes the gold standard for prognosis assessment. We propose a systematic investigation of the value of Artificial Intelligence (AI) to leverage [⁸²Rb] Silicon PhotoMultiplier (SiPM) PET MPI for MACE prediction. We establish a general pipeline for AI model validation to assess and compare the performance of global (i.e. average of the entire MPI signal), regional (17 segments), radiomics and Convolutional Neural Network (CNN) models leveraging various MPI signals on a dataset of 234 patients. Results showed that all regional AI models significantly outperformed the global model ($p < 0.001$), where the best AUC of 73.9% (CI: 72.5-75.3) was obtained with a CNN model. A regional AI model based on MBF averages from 17 segments fed to a Logistic Regression (LR) constituted an excellent trade-off between model simplicity and performance, achieving an AUC of 73.4% (CI: 72.3-74.7). A radiomics model based on intensity features revealed that the global average was the least important feature when compared to other aggregations of the MPI signal over the myocardium. We conclude that AI models can allow better personalized prognosis assessment for MACE.

Introduction

Major Adverse Cardiac Events (MACE) commonly designate stroke, Myocardial Infarction (MI) or cardiac death. As of 2023, those events are the most common outcomes of cardiovascular diseases and remain among the leading causes of death across the world¹. Personalized patient prognosis and the prediction of MACE is therefore a challenge of real interest. Well established risk factors include sedentarity, obesity, diabetes mellitus, arterial hypertension, dyslipidemia, smoking history, and family history of cardiovascular disease²⁻⁴. All those clinical features are not accurate enough when employed for patient personalized prediction, since a significant proportion of the population with cardiovascular diseases does not manifest any of those classical risk factors⁵.

Myocardial Perfusion Imaging (MPI) using Single-Photon Emission Computed Tomography (SPECT) is the most frequently performed nuclear cardiology procedure. It is a sensitive tool for the detection, the localization, and the risk stratification of ischemic heart disease, as well as the assessment of Left Ventricular (LV) function, and myocardial viability. Furthermore, MPI based on Magnetic Resonance Imaging (MRI) is used for its anatomic detail, tissue contrast, spatial and temporal resolution, as well as the lack of ionizing radiation⁶. As an alternative, quantitative MPI can be obtained using [⁸²Rb] Positron Emission Tomography (PET) and is nowadays a functional and non-invasive method for assessing the risk of MACE, or to investigate physiological consequences on the organ after a cardiac event⁷⁻¹². It is already known to bring a strong predictive value for MACE prediction¹³ and was proven cost effective¹⁴. This procedure quantifies the Myocardial Blood Flow (MBF) and the

Myocardial Flow Reserve (MFR) in the LV using dynamic [^{82}Rb] PET/CT acquisitions. MBF at stress (sMBF) and rest (rMBF) can be accurately quantified in ml/min/g¹⁵. The MFR constitutes the ratio of MBF during maximal coronary vasodilatation to resting MBF and is therefore impacted by both rest and stress flow. Thus, it represents the relative reserve of the coronary circulation. Both MBF and MFR are then mapped onto a Polar Map (PM) for visualization. PM visualization yields a controlled positioning and parcellation of all subregions of the LV across patients and acquisitions, based on the 17-segment model of the American Heart Association (AHA)¹⁶.

Various alterations of the MBF were identified to be associated with an increased risk of MACE. In particular, physiological expectations of global MBF impairment are related to multi-vessel epicardial disease or microcirculatory dysfunction. Global sMBF (i.e. average across the entire PM as a measure of the total perfusion of the LV) is known to be predictive for MACE¹³. More isolated alterations of the myocardial perfusion could be related to small defects caused by epicardial coronary artery disease¹⁷. To further investigate this hypothesis, Gould et al. introduced the Myocardial Flow Capacity (MFC) combining sMBF and MFR in a two dimensional representation¹⁷ for categorizing MACE risk.

Nevertheless, the wealth and complexity of the information contained in PMs can be difficult to fully leverage with the naked eye (e.g. micro lesions or subtle patterns) and are subject to inter- and intra- observer variability with semiquantitative visual assessment alone¹⁸. Better individualized risk estimates, with epidemiological effectiveness but also cost efficiency are needed. Artificial Intelligence (AI) has the potential to fully exploit the information provided by PET MPI and dramatically enhance the utility of this powerful modality. Machine Learning (ML)-based PM interpretation already surpassed the clinical interpretation of SPECT MPI for MACE prediction¹⁹. The incorporation of AI techniques to standardize and automate processing of PET MPI could further improve cardiovascular risk stratification^{20,21} to noninvasively support clinical decision for using coronary revascularization⁹ and guide the clinical management of patients with suspected coronary artery disease. This would provide an accurate and systematic assessment of tissue perfusion hemodynamics in a one-stop-shop.

In the specific context of MPI assessed via nuclear medicine and molecular imaging, the very first AI studies originated in the 90s. A large body of literature focused on SPECT due to its large usage and availability²². In 1995, Hamilton et al. used a three-layered Multi-Layer Perceptron (MLP) fed by vectorized 24 territorial values of PMs from [^{201}Tl] SPECT to classify normal and abnormal regions with an AUC (Area Under the receiver operating characteristic Curve) of 96%²³. In 2020, Slomka et al. established the REgistry of Fast Myocardial Perfusion Imaging with NExt generation SPECT/CT (REFINE SPECT)²⁴, an outstanding data resource with MACE as the primary endpoint. It includes more than 20,000 patients from nine centers with [$^{99\text{m}}\text{Tc}$] SPECT along with extensive additional relevant parameters concerning patient data, ECG and treatment. Using the REFINE SPECT database, the prognostic performance of the semi-quantitative assessment of SPECT PMs via stress Total Perfusion Deficit (TPD)²⁵ was found to be superior to visual assessment for predicting MACE²⁶. Betancur et al. used LogitBoost²⁷, a ML approach based on clinical and semi-quantitative imaging variables from [$^{99\text{m}}\text{Tc}$] SPECT (e.g. TPD) to predict MACE (3-year risk), which outperformed visual and semi-quantitative assessments²⁸.

Radiomics aims at extracting large collection of quantitative image measurements describing intensity, shape and texture of regions of interest. Its value was also investigated in the context of nuclear cardiology, with prediction of dilated cardiomyopathy in SPECT MPI²⁹, contraction patterns in gated SPECT MPI³⁰, normal/abnormal and low-risk/high-risk classification in SPECT³¹ as well as detection of diffusely impaired myocardial perfusion in [^{13}N] ammonia PET MPI³². The stability of radiomics features across SPECT scanners was investigated in³³.

Deep Learning (DL) is a subcategory of ML that can directly use images as input. It can deduce and extract optimal image features for the task at hand, obviating the need to handcraft specific features as it is the case for radiomics ML models for instance. [$^{99\text{m}}\text{Tc}$] SPECT-based DL using a Convolutional Neural Network (CNN) was found to be marginally more predictive of obstructive Coronary Artery Disease (CAD) when compared to semi-quantitative assessment via stress TPD³⁴. The accuracy of a CNN based on multiple channels input including sMBF, LV wall motion and wall thickening maps obtained from [$^{99\text{m}}\text{Tc}$] SPECT/CT for the diagnosis of obstructive CAD was assessed by the REFINE SPECT study³⁵. Apostolopoulos et al. used CNNs with [$^{99\text{m}}\text{Tc}$] SPECT PMs to classify 216 patients with either flow-limiting- or no- CAD, which reach performance on par with physicians^{36,37}. Spier et al. compared various DL approaches to classify between normal and abnormal [$^{99\text{m}}\text{Tc}$] SPECT stress and rest PMs³⁸. In particular, they compare MLPs, CNNs on flattened PMs (i.e. either reshaping the PM pixel organization to a square grid, or padding the corners with zeros) and Graph CNNs (GCNNs). A large performance gain is observed with GCNNs when compared to CNNs and MLPs, suggesting that adequate consideration of PM geometry is crucial.

First studies on the relevance of AI based on PET in coronary artery disease were recently published³⁹. Juarez-Orozco et al. used LogitBoost ML with 4 demographic, 8 clinical, and 9 functional variables from [^{13}N] ammonia PET/CT to predict myocardial ischemia and MACE⁴⁰. Wang et al.⁴¹ compared the performance of a Support Vector Machine (SVM) ML model based on 6 MPI variables derived from both [^{13}N] ammonia and [^{18}F]FDG PET/CT to predict the presence of vascular stenosis in patients with suspected obstructive CAD and achieved an AUC of 68%. A multi-task ML based on [^{13}N] ammonia PET/CT was developed by Yeung et al. to identify impaired MFR as well as cardiovascular risk factors⁴². Kwiecinski et al. used ML based on [^{18}F] sodium fluoride PET and quantitative plaque analysis on CT angiography to predict the risk of MI⁴³.

To date, the potential of AI methods to leverage the wealth of [^{82}Rb] PET has been little explored. Wang et al. used a MLP to increase MPI quality⁴⁴ and DL was used to improve motion correction in⁴⁵. Singh et al.⁴⁶ used DL based on multi-channel [^{82}Rb] PET PMs (i.e. stress and rest MBF, MFR, and spill-over fraction computed using QPET⁴⁷) augmented with sex, shape indices⁴⁸ and LV end-systolic and diastolic volumes to predict all-cause mortality. They reported significantly improved mortality prediction when compared to established measures of ischemia.

This study aims to provide a systematic exploration of the prognostic value of [^{82}Rb] PET-based MPI analysis using AI for MACE prediction. To this end, we systematically evaluate and compare standard AI approaches from simple Logistic Regression (LR) based on handcrafted features (e.g. global or regional intensity or radiomics with and without texture features) to CNNs, as well as their combined performance with clinical data. In particular, we address the following research questions when considering MACE prediction: (i) do regional AI models outperform global sMBF (i.e. the traditional quantitative assessment¹³)? (ii) can spatial pattern analysis (i.e. radiomics texture or CNN) improve intensity aggregation (i.e. radiomics first order statistics)? (iii) can we combine MPI and clinical data to improve MACE prediction?

1 Methods

This section systematically details the different steps followed across the study, starting with the descriptions of study population, images acquisition protocols, and data format. We then present all AI models considered and the evaluation pipeline for the classification task of predicting MACE based on [^{82}Rb] PET.

Studied population

Participants with suspected myocardial ischemia were enrolled to undergo [^{82}Rb] cardiac Silicon PhotoMultiplier (SiPM) PET/CT at the Lausanne University Hospital between June 2018 and June 2019¹³. All their cardiovascular risk factors and medication use were determined at time of PET imaging. The Local Ethics Committee, Commission cantonale d'Éthique de la Recherche sur l'être humain VauD (CER-VD) approved this study protocol (#PB_2017 – 00634), and all participants gave written informed consent prior to inclusion. All methods were performed in accordance with the relevant guidelines and regulations. The dataset used for this study consists of MBF measurements at rest and under stress for 234 patients, along with a set of 19 clinical features for each patient (see Supplementary Material A). Among this cohort, there were 187 patients with no observed MACE event in the follow-up days and 47 patients labeled as having a MACE event. The median number of days between PET imaging and first MACE or last news was 652 (interquartile range 559-751). The following events were considered as MACE: cardiac death, delayed revascularization (more than 6 months post-PET/CT), MI, hospitalization for congestive heart failure, or de novo stable angina. MI was defined by clinical presentation compatible with an ST-segment elevation MI or a non-ST-elevation MI, angiographic confirmation of coronary artery disease in the appropriate territory, and an elevated troponin level. Baseline clinical characteristic are detailed in Table 1. More details concerning data acquisition can be found in Dietz et al.¹³.

Imaging protocol

All subjects underwent a rest and adenosine or regadenoson stress SiPM PET/CT scan using a single scanner (Biograph Vision 600, Siemens Medical Solutions, Knoxville, USA). They fasted for 6 hours and avoided caffeine 24 hours before the test. At rest, a 15–25 seconds (s) i.v. infusion of 5 MBq/Kg of [^{82}Rb] was injected with an automatic infusion system (Ruby-Fill® generator and [^{82}Rb] elution system [v3], Jubilant DraxImage, Kirkland, QC, Canada). 3D dynamic PET images were acquired starting with the infusion over 6 minutes (min) 6 s (12×8 , 5×12 , 1×30 , 1×60 , and 1×120 s). For stress acquisitions, adenosine (140 mg/min/kg over 6 minutes) or regadenoson (400 μg over 10-20 seconds) were administered i.v., followed by [^{82}Rb] infusion at 2 min after adenosine and directly after regadenoson administration. The stress protocol was then similar to the rest acquisition. An accompanying low-dose CT (100 keV, 16 mAs) transmission scan was acquired for attenuation correction. An ordered subset expectation maximization algorithm was used for image reconstruction (4 iterations, 5 subsets, 4.0 mm FWHM gaussian post-filter, 220×220 pixel matrix size). Blood pressure, heart rate, and a 12-channel ElectroCardioGram (ECG) were collected during the image acquisitions. Imaging characteristics are presented in Table 2.

Quantitative myocardial perfusion assessment and data format

Perfusion was quantitatively assessed via MBF in ml/g/min at rest and stress, using the FlowQuant v2.7 software (Ottawa, Ontario, Canada) based on a 1-tissue compartment model and flow-dependent extraction correction⁴⁹.

We established the rate-pressure product adjusted rest MBF and the resultant MFR in consideration of variations in resting flow due to differing haemodynamic conditions⁵⁰. This was achieved by multiplying the rest MBF by 8500 mmHg/min and then dividing it by the rate-pressure product (resting heart rate multiplied by resting systolic blood pressure). A dual spill-over correction⁵¹ as well as global partial-volume recovery correction and motion correction⁵² were systematically applied to reduce

	Overall population (n=234)	MACE (n=47)	No MACE (n=187)	p-value
Age, years	72 [61-78]	73 [68-79]	71 [60-77.5]	0.105
Male sex	153 (65%)	36 (77%)	117 (63%)	0.081
Body mass index, kg/m ²	31 [28-36]	32 [28.5-35.5]	31 [28-36]	0.78
Cardiovascular risk factors				
Hypertension	171 (73%)	38 (81%)	133 (71%)	0.18
Current or former smoker	106 (45%)	26 (55%)	80 (43%)	0.12
Dyslipidemia	159 (68%)	34 (72%)	125 (67%)	0.47
Diabetes	85 (36%)	20 (43%)	65 (35%)	0.32
Insulin-requiring diabetes	35 (15%)	10 (21%)	25 (13%)	0.14
Known CAD	126 (54%)	34 (72%)	92 (49%)	0.004
History of MI	100 (43%)	28 (60%)	72 (39%)	0.01
Medications				
Aspirin	135 (58%)	28 (60%)	107 (57%)	0.35
Beta-blockers	144 (62%)	33 (70%)	111 (59%)	0.15
ACE inhibitors/ARB	135 (58%)	32 (68%)	103 (55%)	0.15
Diuretics	82 (35%)	27 (57%)	55 (29%)	0.0002
Nitroglycerine therapy	24 (10%)	6 (13%)	18 (10%)	0.61
Lipid-lowering agent	154 (66%)	32 (68%)	122 (65%)	0.71

Table 1. Baseline clinical characteristics¹³. Reported values are either median [interquartile range] or *n* (%). Angiotensin-Converting Enzyme (ACE); Angiotensin Receptor Blocker (ARB).

	Overall population (n=234)	MACE (n=47)	No MACE (n=187)	p-value
PET pharmacological stress agent				
Adenosine	204 (87%)	40 (85%)	164 (88%)	0.56
Hemodynamics during PET/CT				
Rest-HR, bpm	70 [61-78]	71 [61-81.5]	69 [61-75.5]	0.12
Stress-HR, bpm	83 [74-95]	82 [75-99]	85 [74-94.5]	0.72
Rest-SBP, mmHg	136 ± 23	134 ± 24	136 ± 22	0.6
Stress-SBP, mmHg	120 [104-137]	118 [101-131]	120 [107.5-138]	0.07
Rest-DBP, mmHg	71 ± 12	69 ± 11	71 ± 12.5	0.3
Stress-DBP, mmHg	61 [54-70]	60 [52-63.5]	62 [55-71]	0.11
Rest-RPP > 8500 mmHg/min	153 (65%)	29 (62%)	124 (66%)	0.60
⁸² Rb quantitative imaging				
Global rMBF, mL/min/g	0.82 [0.65-1.06]	0.72 [0.51-0.925]	0.75 [0.59-0.97]	0.34
Global sMBF, mL/min/g	1.96 [1.32-2.71]	1.5 [1.08-1.87]	2.16 [1.54-2.88]	< 0.0001
Global sMBF < 1.94 mL/min/g	113 (48%)	37 (79%)	76 (41%)	< 0.0001
Global MFR	2.39 [1.72-3.0]	1.75 [1.395-2.47]	2.49 [1.93-3.1]	< 0.0001
Global MFR < 1.98	78 (33%)	29 (62%)	49 (26%)	< 0.0001

Table 2. Myocardial perfusion imaging characteristics¹³. Reported values are either median [interquartile range] or *n* (%). Diastolic Blood Pressure (DBP); Heart Rate (HR); Systolic Blood Pressure (SBP); Rate-Pressure Product (RPP = HR × SBP).

the potential spillover in image-derived blood activity curves. Semi-automated segmentation of the myocardium was performed using FlowQuant.

The MBF measurements are provided in the form of a polar-pixelized matrix stored in 24×36 matrices of nonnegative float numbers. The last 8 rows of those matrices were cut off to match with the set of values commonly used to produce the standard PMs representing the perfusion of the LV (see the mapping between the MBF matrix and the MBF PMs according to the 17-segment model in Fig. 1). A measurement matrix is denoted by $S = (S[m, n])_{1 \leq m \leq 16, 1 \leq n \leq 36} \in \mathbb{R}^{16 \times 36}$. Each value $S[m, n]$ represents the perfusion signal at a given polar pixel, where m characterizes the radius to the pole and n the angular position of the polar pixel. Note that the matrix has a periodic structure in the sense that the columns 1 and 36 are spatial neighbors in the PM. MBF values were acquired under stress (sMBF) and at rest (rMBF), leading to two measurement matrices

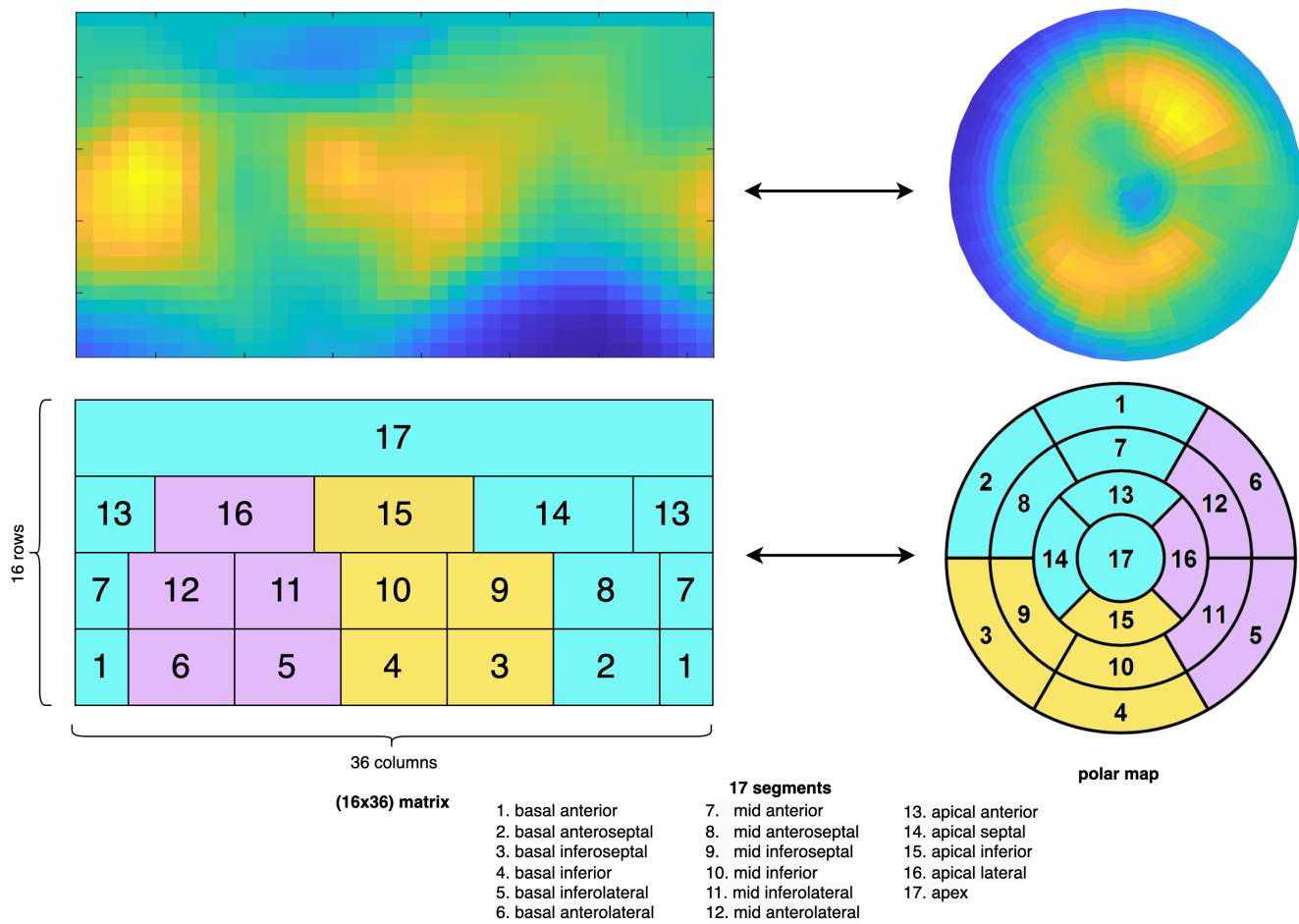


Figure 1. Mapping between 16×36 measurement matrix (left) and Polar Map (PM) (right) representation. Top row: Example of MBF measurements in matrix and PM representation. Bottom row: PM representation of the myocardial perfusion with 17-segment AHA parcellation. 3-segment parcellation is depicted with yellow (inferior wall), purple (lateral wall) and cyan (anterior wall) colors. The mapping between the 16×36 MBF matrix S (left) and the MBF PM (right) is depicted.

S_{sMBF} and S_{rMBF} . For all image analysis approaches, we chose to work with the tabular format $S[m, n]$ instead of the PM images. This allowed to avoid any kind of data transformation such as spatial resampling or intensity binning needed to construct the PM image, as well as empty corners around the PM. This tabular representation was also used with CNNs in Spier et al.³⁸.

Three candidate MPI signals were considered for MACE prediction based on Dietz et al.¹³:

- sMBF: MBF values measured under stress S_{sMBF} in ml/min/g,
- MFR values, obtained by taking the ratio between MBF values under stress and at rest for every element of the MBF matrix; that is $S_{\text{MFR}}[n, m] = \frac{S_{\text{sMBF}}[n, m]}{S_{\text{rMBF}}[n, m]}$ and
- MFC radius signal defined as $S_{\text{MFC}}[n, m] = \sqrt{S_{\text{sMBF}}[n, m]^2 + S_{\text{MFR}}[n, m]^2}$ based on the initial findings of Dietz et al.⁵³.

The first two signals (sMBF and MFR) are established methods for the quantitative assessment of myocardial perfusion and known to be predictive for MACE¹³. We introduce MFC radius as a simple geometric method to evaluate MFC at a pixel level to generate a "capacity" image⁵³. We recall that MFC is able to leverage information from both sMBF and MFR¹⁷ and was previously reported to be predictive of MACE¹³. In what follows, we denote by S the measurement matrix, which is one of the three MPI signals introduced above. We also denote by $\mathbf{y} \in \mathbb{R}^n$ the feature vector, where n is the number of considered features for the corresponding model. The various feature vectors \mathbf{y} considered are systematically detailed in the models' descriptions below (i.e. segment-based or radiomics).

Segment-based models

The first two models that we evaluated were based on two distinct parcellations of the considered PMs, hence of the 16×36 MBF matrices S (see Fig. 1). The global model of the PM corresponds to a global average over the matrix (i.e. over the entire LV), leading to the single-valued measure $y = \frac{1}{16 \times 36} \sum_{1 \leq m \leq 36, 1 \leq n \leq 16} S[m, n]$. We then learn a straightforward LR model taking the global average value of the matrix as unique feature. The 1-segment model was already reported to be predictive for MACE by Dietz et al.¹³ and will serve as our baseline. The second model is based on the standard 17-segment AHA model¹⁶ and corresponds to regional averages, yielding a vector $\mathbf{y} \in \mathbb{R}^{17}$ of 17 features over each delimited region as in Fig. 1. All 17 regional features were standardized based on their respective averages and standard deviations of training samples.

Radiomics models

Two radiomics models were considered. For the first radiomics model (referred to as "radiomics all"), a standard set of radiomics features was included⁵⁴, excluding shape features as not relevant for analysing MPI. All features were computed over the entire measurement matrix S built based on one of the three parametric maps listed above. Using regional masks (e.g. 17 segments) could potentially limit the risk of aggregating radiomics features over too large regions and therefore discarding important local information⁵⁵, but were not considered to limit the number of radiomics features and risk of overfit. We used intensity (18 features) as well as texture features including Gray Level Cooccurrence Matrix (GLCM, 24 features), Gray Level Run Length Matrix (GLRLM, 16 features), Gray Level Size Zone Matrix (GLSZM, 16 features), Neighbouring Gray Tone Difference Matrix (NGTDM, 5 features), and Gray Level Dependence Matrix (GLDM, 14 features), for a total of 93 features yielding a vector $\mathbf{y} \in \mathbb{R}^{93}$. The pyradiomics library was used⁵⁴ with default settings and IBSI compatible in terms of feature definition⁵⁶. We set the *binWidth* separately for each of the MPI signal considered (i.e. sMBF, MFR and MFC) as the max range of the values across the whole data divided by 16. All features were standardized based on their respective averages and standard deviations of training samples. The second radiomics model (referred to as "radiomics intensity") was only based on the 18 intensity features and all texture features were left out, yielding a vector $\mathbf{y} \in \mathbb{R}^{18}$. The two abovementioned radiomics models were used to investigate the specific value of spatial pattern analysis via texture features when compared to simpler intensity measurements.

Machine learning

For all handcrafted models, i.e. global, 17-segment, radiomics all and radiomics intensity models, the respective feature vectors \mathbf{y} fed into either LR models for classification tasks or Cox Proportional Hazard (PH)⁵⁷ regression models for predicting risks that are concordant with time-to-events. For either the LR or the Cox PH models, an ElasticNet penalty combining L_1 and L_2 regularization⁵⁸ was optimized by 5-fold cross-validated grid-search over a range of L_1 _ratio $\in [0.50, 0.85, 0.9, 0.92, 0.95, 0.97, 0.98, 0.99, 0.995, 0.999, 1.0]$. This range was chosen to promote L_1 penalty and thus the sparsity of the model. The optimized score was the AUC or the Concordance (C)-index⁵⁹ to determine the best L_1 _ratio over the training set. For classification models, we represent the outcome *no MACE* (no event) versus *MACE* (event) by the output variable $z \in \{0, 1\}$, so that we learn the parameters of the different LR on the (\mathbf{y}, z) relations. For Cox PH models, time-to-event, events and censoring outcome data are provided to learn the prediction of hazard scores z . The LR and Cox PH models were chosen for their simplicity and well-established methods for radiomics studies. All computational operations were performed with Python v3.9.13, and mainly using scikit-learn v1.0.2 and TensorFlow v2.9.1 libraries.

Late fusion with clinical features

In order to investigate the predictive performance when combining clinical and imaging information, a late fusion of each distinct image model with a LR model based on 19 clinical features (see Supplementary Material A), i.e. $\mathbf{y} \in \mathbb{R}^{19}$ was implemented. To do so, the decision function was simply based on an unweighted average of the prediction scores (probabilities) of the considered image and clinical models. The clinical LR model was trained on the exact same splits as the image-based models (see Fig. 3). Categorical variables were transformed using one-hot encoding and all 19 clinical features were standardized based on their respective averages and standard deviations of training samples.

AI models based on Convolutional Neural Networks

We explored the ability of CNNs to learn and capture more complex spatial patterns (see Fig. 2) as compared to the LR-based models. The measurement matrices S were used as input. We compared two shallow architectures with increasing depths (CNN 1 versus CNN 2). Deeper and pre-trained CNNs (i.e. Resnet50) have been subject to a preliminary evaluation which has not led to any improvement in results. Resnet50 is too complex (i.e. too many layers and parameters) with too large receptive fields to analyse the relatively small input images (16×36) considered in this work.

For all CNN models, an extra step consisting of data augmentation was included in the general pipeline, which was carried out after each train-test split. Since the acquisition of the images is very controlled, we only introduced minor transformations on the PMs simulating the effect of annotation uncertainty. In order to do so, the size of the training set was quadrupled using two types of transformations on the measurement matrices S , corresponding to small variations in the segmentation of the LV. The first transformation was to apply simple shifts of the columns of the matrix (on the right and on the left), corresponding to rotations of the PM of ± 10 degrees (in both directions). Formally, we define the new matrix with the shift on the right by S' with $S'_{ij} = S_{i(j-1)}$ for $j = 2, \dots, 36$ and $S'_{i1} = S_{i36}$. Similarly, we define the new matrix with the shift on the left by S'' with $S''_{ij} = S_{i(j+1)}$ for $j = 1, \dots, 35$ and $S''_{i36} = S_{i1}$. For the second transformation, the last 3 rows of the matrix (corresponding to the outer part of the PM) were erased, and the remaining 13 rows were stretched back to 16 rows using linear interpolation. This simulates variations in the adjustment of the segmentation of the LV on the PET-scan.

For the simplest CNN, referred to as “CNN 1”, we input the matrices and used a convolution layer with 64 filters of size 5×5 , before the Global Average Pooling (GAP) layer. The architecture ended with a Fully Connected (FC) layer including 256 hidden neurons and 2 output neurons. The Rectified Linear Unit (ReLU) activation function was used for the convolutions and FC layers. A softmax activation function was used for the FC output layer (see CNN 1 in Fig. 2). For training the CNNs, we used a L_1 - L_2 regularizer, an Adam optimizer, and a categorical cross-entropy as the loss function. 50 epochs were used for training.

A slightly deeper architecture was considered as used by Spier et al. to analyze SPECT-based MPI in³⁸ (see “CNN 2” in Fig. 2). Similarly to the CNN 1 model, the MBF matrices were used as input followed by two successive convolution layers. The first one included 64 filters of size 5×5 , and the second one had 128 filters of size 3×3 . Each of them used a ReLU activation function, and was followed by a (2×2) -max-pooling layer. A flattening layer was used before the FC layer including 256 hidden neurons and 2 output neurons. ReLU and softmax activation functions were used similarly to the previous model.

General pipeline for AI model validation

In order to systematically compare the different models and assess performance variability, we followed the same procedure for each model family. We repeated 100 random stratified splits of the data into a training and a test set, with a 75-25% proportion (i.e. 175 training cases, including 35 positive, i.e. events, and 140 negative ones). Seeds were kept to ensure the use of same splits for all models. Within each split, the minority class was oversampled in the training set by triplicating every positive case, i.e. events, increasing the training size up to 245 cases (counting 105 positive and 140 negative cases). This was done to address the imbalance of our data, whose positive outcomes only represent a fifth of the total sample. For each tested model, the splits and the shuffling of the training set were exactly the same. The LR, Cox PH and CNN models were then trained, and their performances measured and stored for each split. For LR and CNN models, the AUC of the considered model was computed for each split, and the Youden index was used to determine the optimal cutoff value providing the optimal trade-off between sensitivity and specificity. We finally report the test average performances measured across the 100 splits. Confidence Intervals (CI) and tests for the average were built from 1000 bootstraps⁶⁰ to compare all approaches. An overview of the general training and validation pipeline is depicted in Fig. 3.

2 Results

The demographics and MBF values of the included patients are detailed in Tables 1 and 2 of Dietz et al.¹³, respectively. Table 3 and 4 summarize all the results for classification and Cox PH models, respectively when using sMBF, MFR and MFC radius. The LR model based on the 19 clinical features only achieves an AUC of 63.0% [61.5, 64.3], with a sensitivity of 70.6% and a specificity of 62.4%. The Cox PH model based on the clinical features only achieves a C-index 0.61 [0.60, 0.63]. Fig. 4

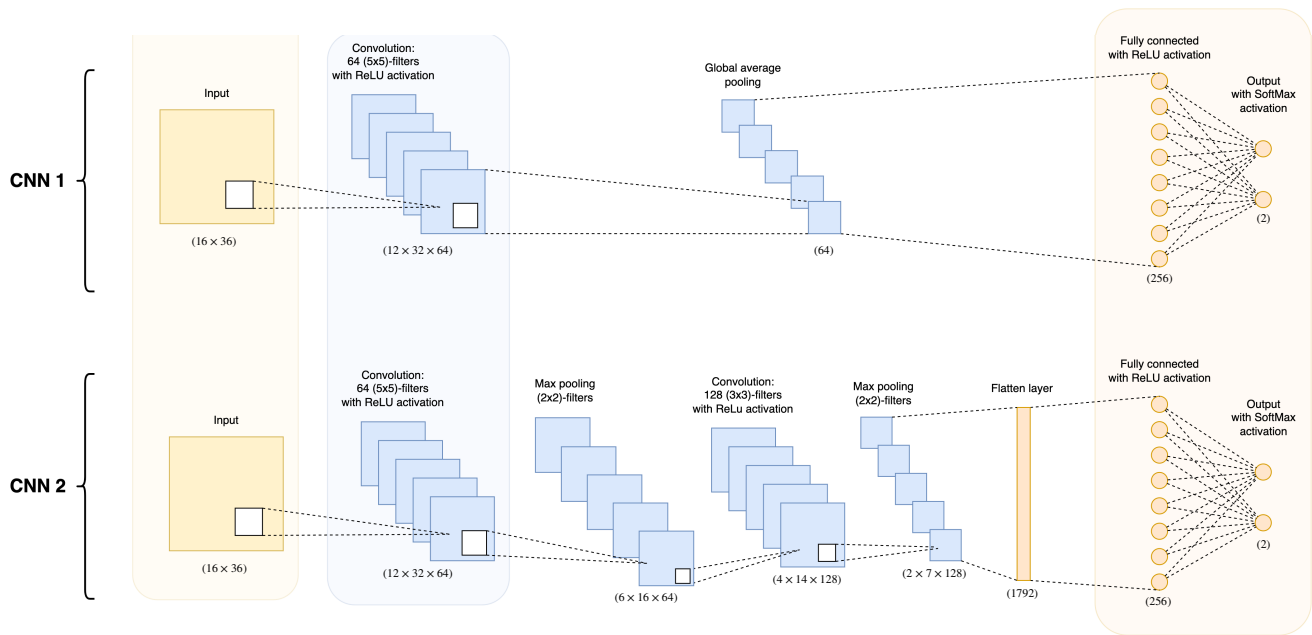


Figure 2. CNN architectures. CNN 2 was proposed by Spier et al. for MPI analysis based on the MBF matrix S^{38} . The number of trainable parameters for CNN 1 and CNN 2 are 19k and 535k, respectively.

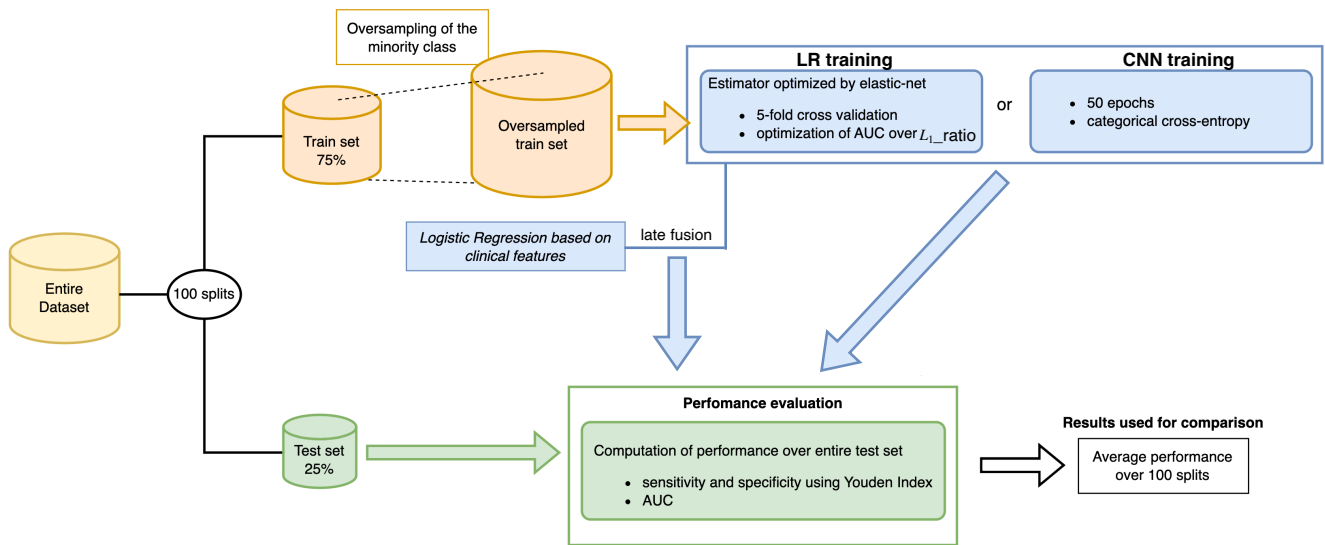


Figure 3. General pipeline for AI model validation.

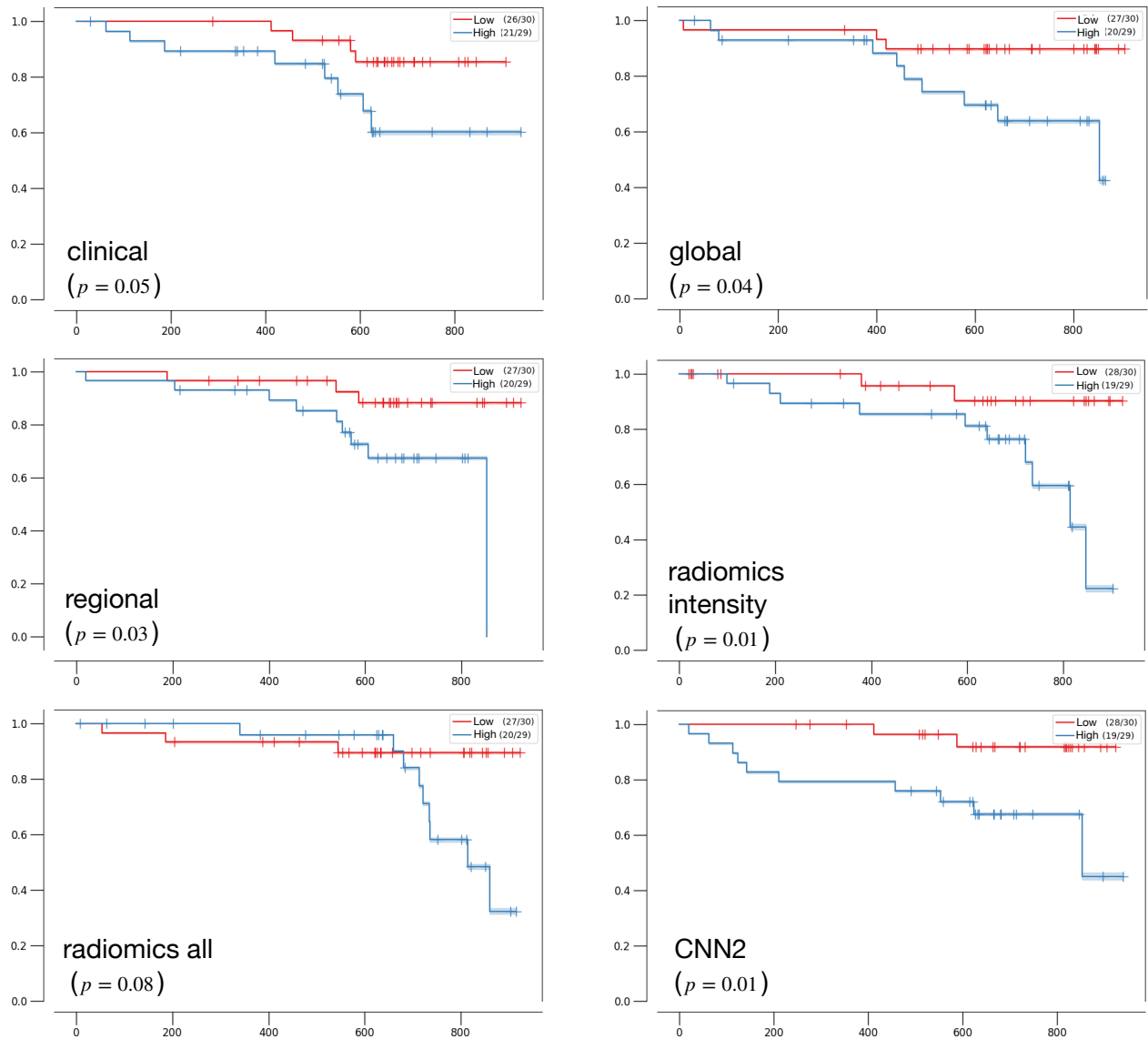


Figure 4. Kaplan-Meier analysis of the models. Low versus high risk groups are split based on the median score. The number of patients per group and censoring are indicated in brackets.

compares Kaplan-Meier curves and log-rank tests based on the probability output scores of the clinical, global and CNN 2 classification models. The low versus high risk groups are based on the median of the probability output scores for each model. Fig. 5, reports the set of features selected by ElasticNet for the radiomics intensity model across all 100 split repetitions (see Fig. 3).

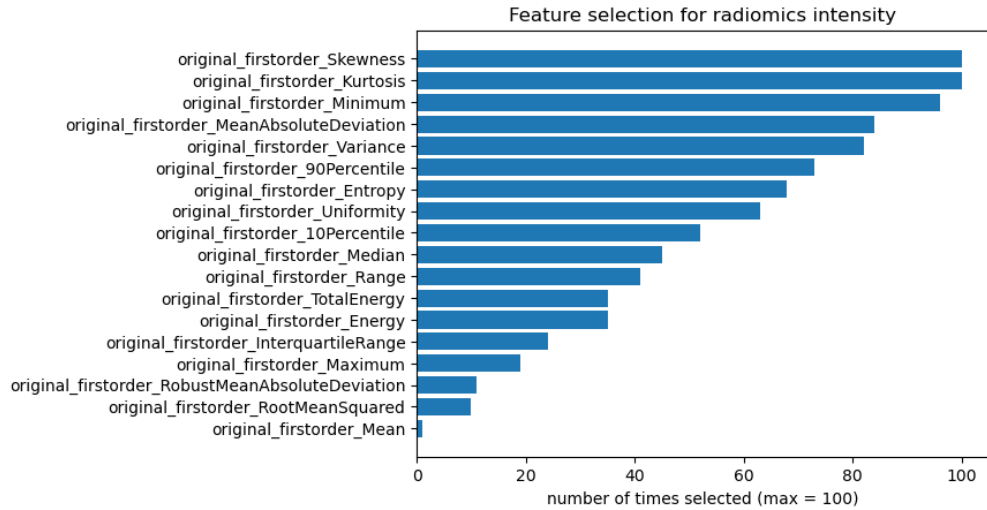


Figure 5. Features selection count for the MFC radius radiomics intensity model over the 100 repeated splits (see Fig. 3). Maximum number of times a feature can be selected is 100, i.e. at each split. The global average of the PM, referred to as “original_firstorder_Mean” is the one chosen least often.

model	# feat.	sMBF			MFR			MFC radius		
		AUC	sens.	spec.	AUC	sens.	spec.	AUC	sens.	spec.
global	1	68.3 [67.0, 69.7]	80.4	62.4	69.7 [68.4, 71.2]	72.5	69.4	70.5 [69.4, 71.8]	77.4	66.0
regional	17	71.2 [69.6, 72.4]	78.1	65.3	71.6 [70.4, 72.9]	74.3	69.8	73.4 [72.3, 74.7]	74.8	72.0
radiomics all	93	67.4 [66.0, 69.0]	72.3	64.7	68.7 [67.2, 70.4]	73.1	67.5	71.6 [70.0, 73.0]	72.4	71.1
radiomics intens.	18	73.2 [71.6, 74.5]	77.1	70.2	68.4 [66.9, 69.8]	73.3	66.8	73.8 [72.4, 75.1]	76.4	67.7
CNN 1	-	70.0 [68.5, 71.3]	80.1	64.3	70.3 [69.0, 71.8]	74.4	71.1	73.6 [72.4, 74.8]	74.8	71.0
CNN 2	-	71.3 [69.7, 72.8]	81.4	64.9	70.8 [69.4, 72.2]	77.3	66.4	73.9 [72.5, 75.3]	77.5	69.0
clinical + global	20	67.6 [66.3, 68.9]	77.6	61.6	68.5 [67.2, 69.8]	75.2	64.6	69.1 [68.0, 70.3]	79.0	62.8
clinical + regional	36	70.9 [69.4, 72.1]	78.2	65.3	71.8 [70.6, 73.0]	76.9	68.4	72.5 [71.3, 73.8]	76.5	70.0
clin. + rad. all	112	69.1 [67.7, 70.5]	72.9	67.1	70.8 [69.1, 72.3]	78.0	67.1	72.8 [71.2, 74.1]	80.0	67.4
clin. + rad. intens.	37	71.5 [70.2, 72.8]	76.7	65.9	69.9 [68.5, 71.3]	77.9	64.5	72.7 [71.3, 73.8]	80.0	67.7
clin. + CNN 1	-	68.2 [66.8, 69.4]	77.7	62.4	69.3 [68.0, 70.5]	76.7	64.7	70.0 [68.9, 71.3]	78.5	64.3
clin. + CNN 2	-	70.4 [69.2, 71.6]	80.1	63.1	71.2 [70.0, 72.5]	78.5	66.4	72.1 [70.9, 73.3]	78.5	67.7

Table 3. Average test performance (in %) of the models computed across 100 stratified random splits. Bootstrap-based 95% confidence intervals are reported for AUC. A LR model based on the 19 clinical features only achieves an AUC of 63.0% [61.5, 64.3], with a sensitivity of 70.6% and a specificity of 62.4%. The top performance is highlighted in bold for each column.

model	# feat.	sMBF	MFR	MFC radius
global	1	0.65 [0.64, 0.66]	0.68 [0.66, 0.69]	0.68 [0.67, 0.69]
regional	17	0.66 [0.65, 0.68]	0.69 [0.68, 0.71]	0.71 [0.70, 0.72]
radiomics all	93	0.66 [0.65, 0.68]	0.66 [0.64, 0.67]	0.69 [0.68, 0.71]
radiomics intens.	18	0.67 [0.66, 0.69]	0.67 [0.65, 0.68]	0.72 [0.71, 0.73]
clinical + global	20	0.64 [0.63, 0.66]	0.67 [0.65, 0.68]	0.66 [0.64, 0.67]
clinical + regional	36	0.68 [0.67, 0.69]	0.70 [0.68, 0.71]	0.71 [0.70, 0.72]
clin. + rad. all	112	0.66 [0.65, 0.68]	0.67 [0.66, 0.68]	0.69 [0.68, 0.71]
clin. + rad. intens.	37	0.66 [0.65, 0.67]	0.67 [0.66, 0.68]	0.69 [0.68, 0.71]

Table 4. Average test C-indices of the Cox PH models computed across 100 stratified random splits. Bootstrap-based 95% confidence intervals are reported. A Cox PH model based on the clinical features only achieves a C-index 0.61 [0.60, 0.63]. The top performance is highlighted in bold for each column.

3 Discussions

In this study, we investigated the value of AI for analyzing [⁸²Rb] PET-based MPI in order to predict MACE. To this end, we systematically implemented and compared standard AI approaches with a global validation pipeline. In particular, we used LRs and Cox PH models based either on MBF segmental (global or regional based on 17 segment) or radiomics (with and without texture) features, as well as shallow CNNs. We also considered a late fusion of imaging signature (global, local, radiomics or CNN) with a LR model uniquely based on clinical features. The three considered input MPI signals including sMBF, MFR and MFC radius were also compared for all models. The training and the evaluation of all these approaches was carried out with a systematic validation pipeline allowing fair and reproducible comparisons between the approaches. The respective performance of models and MPI signals were evaluated based on AUC, sensitivity, specificity and C-index estimated across 100 random train/test splits of the original data.

We interpret the comprehensive performance comparisons presented in Tables 3 and 4 as follows. First, the use of sMBF or MFR seems to perform equivalently on most of the models, whereas for the MFC radius, the performance is consistently better. When focusing on the models based on MFC radius, we observe a significant improvement in average performance when using the regional information across the 17 segments versus using the global average alone (average AUCs of 73.4% and 70.5%, respectively, $p < 0.001$ and average C-indices of 0.68 and 0.71, respectively, $p < 0.001$). Next, we note that the models including clinical features do not yield higher AUCs or C-indices than models that purely rely on image information. For MFC radius, the late fusion seem to consistently improve sensitivity, at the high cost of decreased specificity and resulting in overall decreased AUC. Thus the added value to combine clinical information with image models remains to be confirmed where more advanced approaches to fuse the two sources of information could be explored. Finally, we observe that the models including texture information such as CNNs and radiomics all do not systematically outperform models based on simpler intensity features (i.e. segment-based and radiomics intensity models). For instance, the performance of the 17-segment regional model (average AUC of 73.4%) was found not to be significantly inferior ($p = 0.33$) to the one of the CNN2 model (average AUC of 73.9%). It indicates that the texture information (i.e. subtle spatial patterns that are mostly invisible to the naked eye) contained in the MBF matrices may not predict MACE better than intensity information alone, and that the ElasticNet models do not fully succeed to select most relevant features in the case of the radiomics all model. This suggests that the spatial characteristics of MBF patterns may not be relevant for MACE prediction. It is worth noting that the texture features were aggregated over the entire MBF matrix, and (3- or 17-) segmental aggregation was not investigated to limit the numbers of features per patient. When analyzing intensity features that are retained in the radiomics intensity model in Fig. 5, it is remarkable that the mean feature from the first-order statistics, being strictly equivalent to the global model, is the feature that is the least often selected. This suggests that the spatial aggregation of the MBF signal requires more sophisticated strategies than global averaging. Skewness, kurtosis and the minimum of the MBF distribution constituted the group of most predictive intensity features. From a physiopathological perspective, the superiority of intensity features over texture features is expected since [⁸²Rb] PET-based MPI provides an absolute quantitative assessment of perfusion capacity and deficits that are related to MACE risk. The Kaplan-Meier analyses reported in Fig. 4 demonstrate that all AI models but radiomics all result in better separability of low versus high risk groups when compared to the global and clinical models.

4 Conclusions

Our study demonstrates for the first time the feasibility of AI-based approaches using dynamic [⁸²Rb] PET/CT data in the assessment of MACE, using quantitative MBF parameters as gold standard. Radiomics intensity and CNN 2 models achieved promising results with observed average AUCs of 73.8% and 73.9%, respectively. Overall we conclude that the regional model constitutes an excellent trade-off between model simplicity and performance, achieving high AUC (73.4%) and best specificity (72%). In comparison, CNN 2 (best AUC of 73.9%) favors sensitivity over specificity. The results are promising, and the moderate diagnostic accuracy achieved by the models could be explained by the following limitations. This study is based on a single center with a relatively limited sample size. It would be of interest to apply this pipeline to a larger database collected across multiple centers. We used normalizing resting flow according to the pressure-rate product, which could inherently reduce the prognostic value of flow reserve to predict MACE risk. While this first study focused on traditional AI methods (i.e. radiomics with LR and simple NNs), more advanced AI methods will be investigated in future work. In particular, the tabular representation (i.e. MFB matrix) used as input to the models suffers from spatial distortion. It is worth noting that using PMs as input would also involve spatial distortions in the apex region. Nevertheless, future work will consider analyzing the 3D PET images or graph-based CNNs as this approach showed to better leverage the spatially distorted content of PMs for MPI abnormality classification in SPECT-based MPI³⁸. Other quantitative measurement from [⁸²Rb] PET can also be considered such as left ventricular ejection fraction⁶¹. Another limitation of this study lies in the imbalance of the data, even though we reduce its impact by addressing the lack of positive cases with the oversampling of the minority class.

References

1. Vaduganathan, M., Mensah, G. A., Turco, J. V., Fuster, V. & Roth, G. A. The global burden of cardiovascular diseases and risk. *J. Am. Coll. Cardiol.* **80**, 2361–2371, DOI: <https://doi.org/10.1016/j.jacc.2022.11.005> (2022).
2. Naghavi, M. *et al.* From vulnerable plaque to vulnerable patient: a call for new definitions and risk assessment strategies: Part i. *Circulation* **108(14)**, 1664–1672, DOI: <https://doi.org/10.1161/01.CIR.0000087480.94275.97> (2003).
3. Mozaffarian, D. *et al.* Heart disease and stroke statistics—2015 update. *Circulation* **131(4)**, e29–e39, DOI: <https://doi.org/10.1161/CIR.000000000000152> (2015).
4. Cupples, L. A., Gagnon, D. R., Wong, N. D., Ostfeld, A. M. & Kannel, W. B. Preexisting cardiovascular conditions and long-term prognosis after initial myocardial infarction: The Framingham Study. *Am. Hear. J.* **125(3)**, 863–872, DOI: [https://doi.org/10.1016/0002-8703\(93\)90182-9](https://doi.org/10.1016/0002-8703(93)90182-9) (1993).
5. Khot, U. N. *et al.* Prevalence of conventional risk factors in patients with coronary heart disease. *JAMA* **290(7)**, 898–904, DOI: <https://doi.org/10.1001/jama.290.7.898> (2003).
6. Wu, K. C. Myocardial perfusion imaging by magnetic resonance imaging. *Curr. Cardiol. Reports* **5**, 63–68, DOI: <https://doi.org/10.1007/s11886-003-0039-7> (2003).
7. Harjulahti, E. *et al.* Global and segmental absolute stress myocardial blood flow in prediction of cardiac events: [15O] water positron emission tomography study. *Eur. J. Nucl. Medicine Mol. Imaging* **48(5)**, 1434–1444, DOI: <https://doi.org/10.1007/s00259-020-05093-2> (2021).
8. Juárez-Orozco, L. E. *et al.* Quantitative myocardial perfusion evaluation with positron emission tomography and the risk of cardiovascular events in patients with coronary artery disease: a systematic review of prognostic studies. *Eur. Hear. J. - Cardiovasc. Imaging* **19(10)**, 1179–1187, DOI: <https://doi.org/10.1093/ehjci/jex331> (2018).
9. Knuuti, J. *et al.* 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes: The Task Force for the diagnosis and management of chronic coronary syndromes of the European Society of Cardiology (ESC). *Eur. Hear. J.* **41**, 407–477, DOI: <https://doi.org/10.1093/eurheartj/ehz425> (2020).
10. N.Weber, B. *et al.* Impaired coronary vasodilator reserve and adverse prognosis in patients with systemic inflammatory disorders. *JACC: Cardiovasc. Imaging* **14(11)**, 2212–2220, DOI: <https://doi.org/10.1016/j.jcmg.2020.12.031> (2021).
11. Zampella, E. *et al.* Combined evaluation of regional coronary artery calcium and myocardial perfusion by 82Rb PET/CT in predicting lesion-related outcome. *Eur. J. Nucl. Medicine Mol. Imaging* **47(7)**, 1698–1704, DOI: <https://doi.org/10.1007/s00259-019-04534-x> (2020).
12. Frey, S. M. *et al.* Artificial intelligence to improve ischemia prediction in rubidium positron emission tomography—a validation study. *EPMA J.* **14**, 631–643, DOI: <https://doi.org/10.1007/s13167-023-00341-5> (2023).
13. Dietz, M. *et al.* Comparison of the prognostic value of impaired stress myocardial blood flow, myocardial flow reserve, and myocardial flow capacity on low-dose Rubidium-82 SiPM PET/CT. *J. Nucl. Cardiol.* DOI: <https://doi.org/10.1007/s12350-022-03155-6> (2022).
14. Mimouni, M. *et al.* Cost-effectiveness of 82-Rubidium PET myocardial perfusion imaging for the diagnosis of myocardial ischemia depending on the prevalence of coronary artery disease. *EJNMMI Res.* **13**, 1–9, DOI: <https://doi.org/10.1186/s13550-023-00954-x> (2023).
15. Prior, J. O. *et al.* Quantification of myocardial blood flow with 82Rb positron emission tomography: Clinical validation with 15O-water. *Eur. J. Nucl. Medicine Mol. Imaging* **39**, 1037–1047, DOI: <https://doi.org/10.1007/s00259-012-2082-3> (2012).
16. Cerqueira, M. D. *et al.* Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart. *Circulation* **105**, 539–542, DOI: <https://doi.org/10.1161/hc0402.102975> (2002).
17. Gould, K. L. *et al.* Anatomic versus physiologic assessment of coronary artery disease: Role of coronary flow reserve, fractional flow reserve, and positron emission tomography imaging in revascularization decision-making. *J. Am. Coll. Cardiol.* **62(18)**, 1639–1653, DOI: <https://doi.org/10.1016/j.jacc.2013.07.076> (2013).
18. Xu, Y. *et al.* Automatic and visual reproducibility of perfusion and function measures for myocardial perfusion SPECT. *J. Nucl. Cardiol.* **17**, 1050, DOI: <https://doi.org/10.1007/s12350-010-9297-0> (2010).
19. Hu, L. H. *et al.* Prognostically safe stress-only single-photon emission computed tomography myocardial perfusion imaging guided by machine learning: report from refine spect. *Eur. Hear. J. - Cardiovasc. Imaging* **22(6)**, 705–714, DOI: <https://doi.org/10.1093/ehjci/jeaa134> (2021).

20. Slomka, P. Future of nuclear cardiology is bright: Promise of cardiac PET/CT and artificial intelligence. *J. Nucl. Cardiol.* 2022 1–3, DOI: <https://doi.org/10.1007/s12350-022-02942-5> (2022).
21. Slomka, P. J. *et al.* Quantitative Clinical Nuclear Cardiology, Part 2: Evolving/Emerging Applications. *J. Nucl. Medicine* **62**, 168–176, DOI: <https://doi.org/10.2967/jnumed.120.242537> (2020).
22. Garcia, E. V. & Piccinelli, M. Preparing for the artificial intelligence revolution in nuclear cardiology. *Nucl. Medicine Mol. Imaging* **57**, 51–60, DOI: <https://doi.org/10.1007/s13139-021-00733-3> (2022).
23. Hamilton, D., Riley, P. J., Miola, U. J. & Amro, A. A. A feed forward neural network for classification of bull's-eye myocardial perfusion images. *Eur. J. Nucl. Medicine* 1995 22:2 **22**, 108–115, DOI: <https://doi.org/10.1007/BF00838939> (1995).
24. Slomka, P. J. *et al.* Rationale and design of the REgistry of Fast Myocardial Perfusion Imaging with NExt generation SPECT (REFINE SPECT). *J. Nucl. Cardiol.* **27**, 1010–1021, DOI: <https://doi.org/10.1007/s12350-018-1326-4> (2020).
25. Nakazato, R. *et al.* Quantitative Upright–Supine High-Speed SPECT Myocardial Perfusion Imaging for Detection of Coronary Artery Disease: Correlation with Invasive Coronary Angiography. *J. Nucl. Medicine* **51**, 1724–1731, DOI: <https://doi.org/10.2967/jnumed.110.078782> (2010).
26. Otaki, Y. *et al.* 5-year prognostic value of quantitative versus visual MPI in subtle perfusion defects: Results from REFINE SPECT. *JACC: Cardiovasc. Imaging* **13**, 774–785, DOI: <https://doi.org/10.1016/j.jcmg.2019.02.028> (2020).
27. Friedman, J., Hastie, T. & Tibshirani, R. Additive logistic regression: a statistical view of boosting. *The Annals Stat.* **28**, 337–407, DOI: <https://doi.org/10.1214/aos/1016218223> (2000).
28. Betancur, J. *et al.* Prognostic value of combined clinical and myocardial perfusion imaging data using machine learning. *JACC: Cardiovasc. Imaging* **11**, 1000–1009, DOI: <https://doi.org/10.1016/j.jcmg.2017.07.024> (2018).
29. Wang, C. *et al.* Texture analysis of SPECT myocardial perfusion provides prognostic value for dilated cardiomyopathy. *J. Nucl. Cardiol.* 2022 1–12, DOI: <https://doi.org/10.1007/s12350-022-03006-4> (2022).
30. Sabouri, M. *et al.* Myocardial perfusion SPECT imaging radiomic features and machine learning algorithms for cardiac contractile pattern recognition. *J. Digit. Imaging* 2022 1–13, DOI: <https://doi.org/10.1007/s10278-022-00705-9> (2022).
31. Amini, M. *et al.* Machine learning-based diagnosis and risk classification of coronary artery disease using myocardial perfusion imaging SPECT: A radiomics study. *Sci. Reports* 2023 13:1 **13**, 1–12, DOI: <https://doi.org/10.1038/s41598-023-42142-w> (2023).
32. Degtiarova, G. *et al.* Radiomics for the detection of diffusely impaired myocardial perfusion: A proof-of-concept study using ¹³N-ammonia positron emission tomography. *J. Nucl. Cardiol.* 2023 1–10, DOI: <https://doi.org/10.1007/s12350-022-03179-y> (2023).
33. Edalat-Javid, M. *et al.* Cardiac SPECT radiomic features repeatability and reproducibility: A multi-scanner phantom study. *J. Nucl. Cardiol.* **28**, 2730–2744, DOI: <https://doi.org/10.1007/s12350-020-02109-0> (2021).
34. Betancur, J. *et al.* Deep learning for prediction of obstructive disease from fast myocardial perfusion SPECT: A multicenter study. *JACC: Cardiovasc. Imaging* **11**, 1654–1663, DOI: <https://doi.org/10.1016/j.jcmg.2018.01.020> (2018).
35. Otaki, Y. *et al.* Clinical deployment of explainable artificial intelligence of SPECT for diagnosis of coronary artery disease. *JACC: Cardiovasc. Imaging* DOI: <https://doi.org/10.1016/j.jcmg.2021.04.030> (2021).
36. Apostolopoulos, I. D., Papathanasiou, N. D., Spyridonidis, T. & Apostolopoulos, D. J. Automatic characterization of myocardial perfusion imaging polar maps employing deep learning and data augmentation. *Hellenic J. Nucl. Medicine* **23**, 125–132, DOI: <https://doi.org/10.1967/s002449912101> (2020).
37. Apostolopoulos, I. D., Apostolopoulos, D. I., Spyridonidis, T. I., Papathanasiou, N. D. & Panayiotakis, G. S. Multi-input deep learning approach for cardiovascular disease diagnosis using myocardial perfusion imaging and clinical data. *Phys. Medica* **84**, 168–177, DOI: <https://doi.org/10.1016/j.ejmp.2021.04.011> (2021).
38. Spier, N., Nekolla, R., S. & C. Classification of polar maps from cardiac perfusion imaging with graph-convolutional neural networks. *Sci. Rep.* **9**, 7569, DOI: <https://doi.org/10.1038/s41598-019-43951-8> (2019).
39. Popescu, C. *et al.* PET-based artificial intelligence applications in cardiac nuclear medicine. *Swiss Med. Wkly.* 2022 :3 **152**, w30123, DOI: <https://doi.org/10.4414/SMW.2022.w30123> (2022).
40. Juarez-Orozco, L. E. *et al.* Machine learning in the integration of simple variables for identifying patients with myocardial ischemia. *J. Nucl. Cardiol.* **27**, 147–155, DOI: <https://doi.org/10.1007/s12350-018-1304-x> (2020).

41. Wang, F. *et al.* Evaluation of the diagnostic value of joint PET myocardial perfusion and metabolic imaging for vascular stenosis in patients with obstructive coronary artery disease. *J. Nucl. Cardiol.* **28**, 3070–3080, DOI: <https://doi.org/10.1007/s12350-020-02160-x> (2021).
42. Yeung, M. W. *et al.* Multi-task deep learning of myocardial blood flow and cardiovascular risk traits from PET myocardial perfusion imaging. *J. Nucl. Cardiol.* **2022** 1–11, DOI: <https://doi.org/10.1007/s12350-022-02920-x> (2022).
43. Kwiecinski, J. *et al.* Machine learning with 18F-Sodium Fluoride PET and quantitative plaque analysis on CT angiography for the future risk of myocardial infarction. *J. Nucl. Medicine* **63**, 158–165, DOI: <https://doi.org/10.2967/jnumed.121.262283> (2022).
44. Wang, X., Yang, B., Moody, J. B., Tang, J. & Wang, X. Improved myocardial perfusion PET imaging using artificial neural networks. *Phys. Medicine & Biol.* **65**, 145010, DOI: <https://doi.org/10.1088/1361-6560/AB8687> (2020).
45. Shi, L. *et al.* Automatic inter-frame patient motion correction for dynamic cardiac PET using deep learning. *IEEE Transactions on Med. Imaging* **40**, 3293–3304, DOI: <https://doi.org/10.1109/TMI.2021.3082578> (2021).
46. Singh, A. *et al.* Deep learning for explainable estimation of mortality risk from myocardial positron emission tomography images. *Circ. Cardiovasc. Imaging* **15**, E014526, DOI: <https://doi.org/10.1161/CIRCIMAGING.122.014526> (2022).
47. DeKemp, R. A. *et al.* Multisoftware reproducibility study of stress and rest myocardial blood flow assessed with 3D dynamic PET/CT and a 1-tissue-compartment model of 82Rb kinetics. *J. Nucl. Medicine* **54**, 571–577, DOI: <https://doi.org/10.2967/JNUMED.112.112219> (2013).
48. Abidov, A. *et al.* Left ventricular shape index assessed by gated stress myocardial perfusion spect: Initial description of a new variable. *J. Nucl. Cardiol.* **2006** 13:5 **13**, 652–659, DOI: <https://doi.org/10.1016/J.NUCLCARD.2006.05.020> (2006).
49. deKemp, R. A. *et al.* Multisoftware reproducibility study of stress and rest myocardial blood flow assessed with 3d dynamic pet/ct and a 1-tissue-compartment model of 82rb kinetics. *J. Nucl. Medicine* **54**, 571–577, DOI: <https://doi.org/10.2967/jnumed.112.112219> (2013).
50. Sdringola, S., Johnson, N. P., Kirkeeide, R. L., Cid, E. & Gould, K. L. Impact of unexpected factors on quantitative myocardial perfusion and coronary flow reserve in young, asymptomatic volunteers. *JACC: Cardiovasc. Imaging* **4**, 402–412, DOI: <https://doi.org/10.1016/J.JCMG.2011.02.008> (2011).
51. Efseaff, M., Klein, R., Ziadi, M. C., Beanlands, R. S. & deKemp, R. A. Short-term repeatability of resting myocardial blood flow measurements using rubidium-82 PET imaging. *J. Nucl. Cardiol.* **19**, 997–1006, DOI: <https://doi.org/10.1007/s12350-012-9600-3> (2012).
52. Hunter, C. R. R. N., Klein, R., Beanlands, R. S. & deKemp, R. A. Patient motion effects on the quantification of regional myocardial blood flow with dynamic PET imaging. *Med. Phys.* **43**, 1829–1840, DOI: <https://doi.org/10.1118/1.4943565> (2016).
53. Dietz, M. *et al.* Comparison of the prognostic value of global and regional myocardial flow capacity radius, myocardial flow reserve, and stress myocardial blood flow using rubidium-82 with SiPM PET/CT. *ResearchSquare* DOI: <https://doi.org/10.21203/RS.3.RS-1122028/V1> (2021).
54. van Griethuysen, J. J. M. *et al.* Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* **77**, e104–e107, DOI: <https://doi.org/10.1158/0008-5472.CAN-17-0339> (2017).
55. Fontaine, P. *et al.* The importance of feature aggregation in radiomics: a head and neck cancer study. *Sci. Reports* **2020** 10:1 **10**, 1–11, DOI: <https://doi.org/10.1038/s41598-020-76310-z> (2020).
56. Zwanenburg, A. *et al.* The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* **295**, 328–338, DOI: <https://doi.org/10.1148/radiol.2020191145> (2020).
57. Simon, N., Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for Cox’s proportional hazards model via coordinate descent. *J. Stat. Softw.* **39**, 1–13, DOI: <https://doi.org/10.18637/jss.v039.i05> (2011).
58. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. Royal Stat. Soc. Ser. B (Statistical Methodol.)* **67**, 301–320, DOI: <https://doi.org/10.1111/j.1467-9868.2005.00503.x> (2005).
59. Harrell, F. E., Lee, K. L. & Mark, D. B. Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat. medicine* **15**, 361–387 (1996).
60. Efron, B. Bootstrap methods: Another look at the jackknife. *The Annals Stat.* **7**, 569–593, DOI: https://doi.org/10.1007/978-1-4612-4380-9_41 (1992).

61. Lertsburapa, K. *et al.* Independent and incremental prognostic value of left ventricular ejection fraction determined by stress gated rubidium 82 PET imaging in patients with known or suspected coronary artery disease. *J. Nucl. Cardiol.* **15**, 745–753, DOI: <https://doi.org/10.1007/BF03007355/METRICS> (2008).

Author contributions statement

Study design: S.B., D.A., C.H.K., J.O.P., A.D. Conducting experiments: S.B., D.A., A.D. Analysis tools development: S.B., D.A., V.A., R.D.K., A.D. Data curation: S.B., M.D., Cohort identification and creation: M.D., J.O.P. Writing of the manuscript: S.B., A.D. All authors reviewed the manuscript and accepted the final version.

Data availability

The datasets generated and/or analysed during the current study are not publicly available as not permitted by the ethics agreement.

Competing interests

R.D.K. receives royalty revenues from Rubidium-82 PET technologies licensed to Jubilant DraxImage and INVIA Medical Solutions. All other authors do not possess any conflict of interest.

Acknowledgments

This work was partially funded by the Swiss National Science Foundation (SNSF) with the projects 205320_219430 and 205320_179069, the Swiss Cancer Research foundation with the project TARGET (KFS-5549-02-2022-R), the Lundin Center for Neuro-oncology at CHUV, and the Hasler Foundation with the project MSxplain number 21042.