THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Clinical phenotypes among patients with normal cardiac perfusion using unsupervised learning

OPEN ACCESS

Clinical phenotypes among patients with normal cardiac perfusion using unsupervised learning: a retrospective observational study

Robert J. H. Miller,[a,b,q] Bryan P. Bednarski,[a,q] Konrad Pieszko,[a] Jacek Kwiecinski,[a,c] Michelle C. Williams,[a,d] Aakash Shanbhag,[a,e] Joanna X. Liang,[a] Cathleen Huang,[a] Tali Sharir,[f,p] M. Timothy Hauser,[g] Sharmila Dorbala,[h] Marcelo F. Di Carli,[h] Mathews B. Fish,[i] Terrence D. Ruddy,[j] Timothy M. Bateman,[k] Andrew J. Einstein,[l] Philipp A. Kaufmann,[m] Edward J. Miller,[n] Albert J. Sinusas,[n] Wanda Acampa,[o] Donghee Han,[a] Damini Dey,[a] Daniel S. Berman,[a] and Piotr J. Slomka[a,*]

[a]Departments of Medicine (Division of Artificial Intelligence in Medicine), Biomedical Sciences, and Imaging, Cedars-Sinai Medical Center, Los Angeles, CA, USA
[b]Department of Cardiac Sciences, University of Calgary and Libin Cardiovascular Institute, Calgary, AB, Canada
[c]Department of Interventional Cardiology and Angiology, Institute of Cardiology, Warsaw, Poland
[d]British Heart Foundation Centre for Cardiovascular Science, University of Edinburgh, Edinburgh, United Kingdom
[e]Signal and Image Processing Institute, Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA, USA
[f]Department of Nuclear Cardiology, Assuta Medical Centers, Tel Aviv, Israel
[g]Department of Nuclear Cardiology, Oklahoma Heart Hospital, Oklahoma City, OK, USA
[h]Division of Nuclear Medicine and Molecular Imaging, Department of Radiology, Brigham and Women's Hospital, Boston, MA, USA
[i]Oregon Heart and Vascular Institute, Sacred Heart Medical Center, Springfield, OR, USA
[j]Division of Cardiology, University of Ottawa Heart Institute, Ottawa, ON, Canada
[k]Cardiovascular Imaging Technologies LLC, Kansas City, MO, USA
[l]Division of Cardiology, Department of Medicine and Department of Radiology, Columbia University Irving Medical Center and New York-Presbyterian Hospital, New York, NY, USA
[m]Department of Nuclear Medicine, Cardiac Imaging, University Hospital Zurich, Zurich, Switzerland
[n]Section of Cardiovascular Medicine, Department of Internal Medicine, Yale University School of Medicine, New Haven, CT, USA
[o]Department of Advanced Biomedical Sciences, University of Naples "Federico II", Naples, Italy
[p]Israel and Ben Gurion University of the Negev, Beer Sheba, Israel

Summary

Background Myocardial perfusion imaging (MPI) is one of the most common cardiac scans and is used for diagnosis of coronary artery disease and assessment of cardiovascular risk. However, the large majority of MPI patients have normal results. We evaluated whether unsupervised machine learning could identify unique phenotypes among patients with normal scans and whether those phenotypes were associated with risk of death or myocardial infarction.

Methods Patients from a large international multicenter MPI registry (10 sites) with normal perfusion by expert visual interpretation were included in this cohort analysis. The training population included 9849 patients, and external testing population 12,528 patients. Unsupervised cluster analysis was performed, with separate training and external testing cohorts, to identify clusters, with four distinct phenotypes. We evaluated the clinical and imaging features of clusters and their associations with death or myocardial infarction.

Findings Patients in Clusters 1 and 2 almost exclusively underwent exercise stress, while patients in Clusters 3 and 4 mostly required pharmacologic stress. In external testing, the risk for Cluster 4 patients (20.2% of population, un-adjusted hazard ratio [HR] 6.17, 95% confidence interval [CI] 4.64–8.20) was higher than the risk associated with pharmacologic stress (HR 3.03, 95% CI 2.53–3.63), or previous myocardial infarction (HR 1.82, 95% CI 1.40–2.36).

Interpretation Unsupervised learning identified four distinct phenotypes of patients with normal perfusion scans, with a significant proportion of patients at very high risk of myocardial infarction or death. Our results suggest a potential role for patient phenotyping to improve risk stratification of patients with normal imaging results.

*Corresponding author. Cedars-Sinai Medical Center, 8700 Beverly Boulevard, Ste. Metro 203, Los Angeles, 90048, CA, USA.
    E-mail address: Piotr.Slomka@cshs.org (P.J. Slomka).
    @Piotr_JSlomka (P.J. Slomka).
[q]Authors contributed equally.

Articles

Keywords: Machine learning; Coronary artery disease; Cluster analysis; Myocardial perfusion

---

Research in context

Evidence before this study

We searched PubMed for articles with the following keywords: "perfusion imaging" AND "unsupervised machine learning" for studies up until August 2023. We identified 9 studies in total, none of which focused on patients with normal myocardial perfusion. Previous studies have evaluated risk associated with specific clinical characteristics. In contrast, unsupervised machine learning has the potential to identify patient phenotypes which reflect multiple patient characteristics.

Added value of this study

We applied an unsupervised machine learning model to identify patient phenotypes among patients with normal myocardial perfusion by expert visual interpretation.

Unsupervised machine identified 4 distinct patient clusters in both internal (n = 9849) and external (n = 12,528) populations. The risk associated with being in the fourth cluster was higher than traditional high-risk groups, such as inability to complete exercise stress or previous myocardial infarction.

Implications of all the available evidence

Our results suggest that machine learning-based phenotyping is potentially a powerful tool for patient classification and risk stratification even when the key imaging result is normal. These findings are relevant for the majority of patients undergoing cardiac perfusion imaging; however, the implications are potentially relevant for a variety of imaging tests.

Introduction

Myocardial perfusion imaging (MPI) is one of the most common heart scans performed for assessment of coronary artery disease (approximately 6 million procedures per year in United States and 15–20 million per year worldwide), second only to echocardiography.[1] This test is frequently used for cardiovascular risk stratification,[2] with the presence of abnormal myocardial perfusion being an established marker of cardiovascular risk.[3,4] However, the prevalence of abnormal myocardial perfusion has decreased dramatically over time.[5,6] In fact, the vast majority (>90% in some cohorts)[5] show normal perfusion.[7] While these patients are generally lower risk, there is still significant heterogeneity among them. The best method to classify risk in this substantial population following MPI is a clinically relevant challenge that physicians face.

When identifying important sub-groups of patients with normal perfusion, physicians have typically relied on co-morbidities such as diabetes[8] or history of coronary artery disease.[9] However, the increase in risk is modest. Alternatively, physicians can look for abnormal imaging findings other than perfusion. For example, reduced left ventricular ejection fraction[10] and transient ischemic dilation of the left ventricle.[11] However, these markers are seen in a minority of studies and therefore are only helpful in identifying a small number of high-risk patients. Identification of clinically relevant phenotypes of patients with normal perfusion remains a critically important and unmet need.

Artificial intelligence techniques have emerged as powerful tools for improving disease diagnosis or risk stratification.[12] These are frequently supervised models, which are trained to predict a specific outcome, such as the presence of coronary disease. Unsupervised machine learning algorithms are not trained to predict specific outcomes. Instead, these algorithms learn and label inherent structures within the data.[12] These techniques can be used to identify clusters, or groups, within a dataset without any preconceptions about the number of groups or which variables are important. These algorithms are not trained to predict specific outcomes, and thus can be developed or updated without waiting for long-term clinical follow-up; however, the resulting phenotypes frequently have important clinical implications. For example, unsupervised learning has been used to identify three distinct phenotypes of patients with heart failure with preserved ejection fraction, and the risks for atherosclerotic and heart failure-related outcomes differ significantly between groups.[13]

We applied an unsupervised machine learning model to identify phenotypes among patients with normal myocardial perfusion by expert visual interpretation on clinically indicated MPI. We then examined

the patient characteristics of these clusters and evaluated their associations with death or myocardial infarction using a large, international, multicenter registry (10 centers) with dedicated training and external testing populations.

Methods

Study design and setting

We included consecutive patients undergoing clinically indicated single photon emission computed tomography MPI between 2009 and 2021 from the international, multicenter, REFINE-SPECT registry.[14] The overall study protocol has been described in detail previously.[14] Data for REFINE-SPECT is collected locally and then de-identified prior to transfer to the core laboratory. At the core laboratory data elements undergo quality assurance processes before integration into the overall registry. Once the data was checked for completion (including the verification of the code match and the check of clinical data against image data), the imaging results and clinical data were merged using Python software version 3.9.3 (Willmington, DE, USA) and integrated into a PostgreSQL database version 14.0 (Philadelphia, PA, USA).

Participants

Patient inclusions and exclusions are shown in Fig. 1 and Supplemental Table S1. After excluding patients without stress imaging (n = 16), follow-up for clinical events (n = 71), and patients with abnormal perfusion by expert visual interpretation (n = 14,836), 22,377 patients were included in the analysis. Abnormal visual interpretation of perfusion was defined as summed stress score >0 or reader diagnosis other than normal if summed stress scores were not available.[3] Visual interpretation was performed at the time of clinical reporting (multiple physicians per site), with knowledge of patient features, including medical history and presenting symptoms. The final patient population was divided a priori by site into an internal cohort for algorithm development (n = 9849 from 4 sites) and an external cohort for testing (n = 12,528 from 6 sites). External testing is critical to evaluate the potential performance of the unsupervised learning model when applied in new sites, previously unseen by the model.

Clinical information

Demographic information and past medical history was collected by each site at the time that MPI was performed. The collected variables that were used in this study are outlined in Supplemental Table S2.

Scan acquisition and interpretation

All scans were acquired according to existing guidelines for clinical purposes, with details available in the supplement. Details of the automated analysis for quantification of perfusion, functional parameters, phase analysis,[15] and shape index[16] are available in the supplement.

Clinical outcomes

The primary objective of the unsupervised machine learning approach was to group data into clusters. We then evaluated the associations with death or non-fatal myocardial infarction to assess the potential importance of the clusters. Details of the method for ascertaining events at each site are available in the Supplemental Materials. All non-fatal myocardial infarction events were adjudicated by experienced physicians using standard criteria.

Unsupervised machine learning

Unsupervised cluster analysis was implemented using Python (Version 3.9.7) with clinical information (23 parameters), stress parameters (12 parameters), image acquisition parameters (5 parameters), and quantitative image analysis parameters (20 parameters, Supplemental Table S2). We also evaluated a model that incorporated only the 20 image analysis parameters. Cardiovascular events and mortality were not used in fitting the unsupervised model. Missing variables were imputed using median imputation for continuous, and mode imputation for categorical variables across the whole population. Variables with >25% missingness in the internal cohort were dropped during preprocessing. The analysis was performed blinded to patient outcomes.

Dimensionality reduction was performed using the non-linear Uniform Manifold Approximation and Projection package (UMAP Learn, Version 0.5.2).[17] Dimensionality reduction improves the performance of cluster analysis by simplifying the input feature space prior to clustering, reducing computation time and noise, while preserving the global data structure (i.e., the relative relationship between patients in the data).[17] UMAP was specifically selected as the primary engine for our unsupervised pipeline as it utilizes non-linear manifold approximation theory to estimate a low-dimensional representation in a more efficient and scalable manner than principal component analysis and t-distributed stochastic neighbor embedding (t-SNE), while retaining a stable model representation that is viable for clinical deployment.[17] Cao et al. demonstrated the robustness of UMAP to embed high dimensional data from cellular biology into a new representation, leading to fewer clusters than t-SNE.[18] Due to its success with higher-dimensional data, such as that available with MPI (clinical, stress, and imaging features available), we utilized UMAP.

In the interest of robust model selection and hyperparameter search, all UMAP models were tested and validated with three different clustering algorithms (hierarchical, k-means, gaussian mixture model; Scikit-
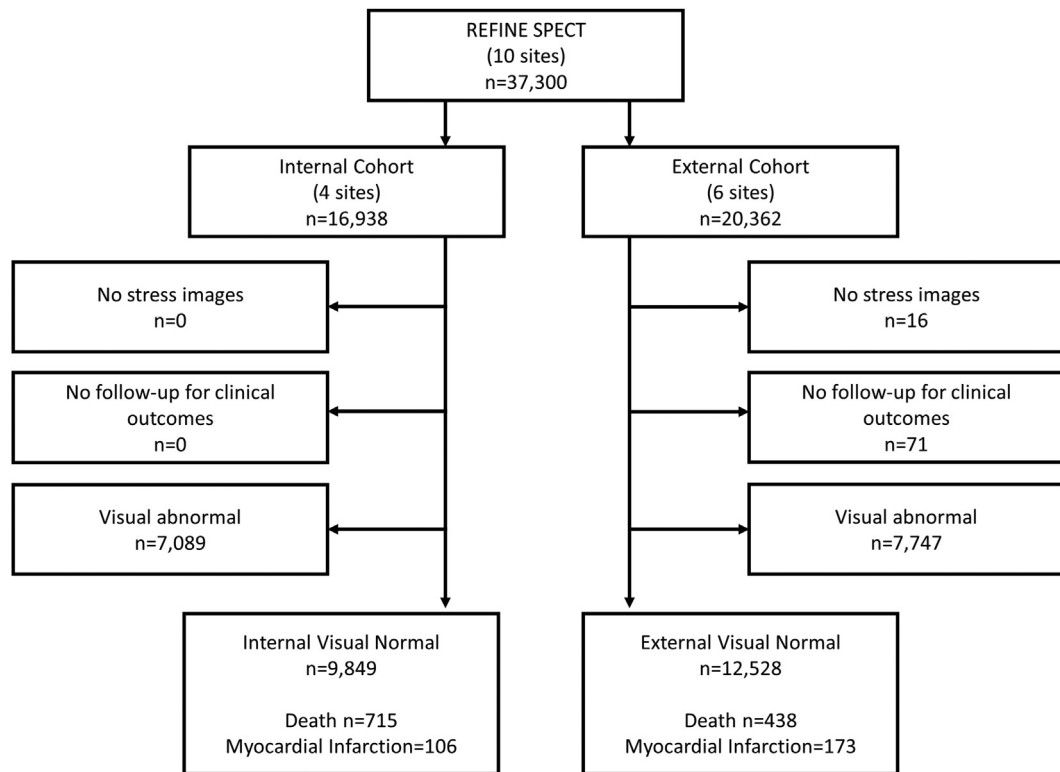
Fig. 1: Population flow diagram outlining patient inclusions and exclusions. Visual normal perfusion was defined as summed stress score of zero or expert reader interpretation of normal if summed stress score was not available. Refer to Supplemental Table S1 for a per-site exclusions summary.

Learn package, Version 1.0.1). The gaussian mixture model algorithm was utilized in the best-performing pipeline from internal cohort training. However, to investigate the impact of clustering algorithm choice, we compared the clustering performance of hierarchical, k-means, gaussian mixture model clustering using the best-performing UMAP model from internal testing (Supplemental Table S3).

A grid search was used to optimize the dimensionality reduction parameters, clustering method, and number of clusters (Supplemental Table S4). The optimal combination of parameters was selected based on the silhouette score. Silhouette scores provide a metric to assess how well clusters are separated, by assessing the separation distance of individuals between different clusters compared to distance within a cluster. Silhouette score is similar to other cluster-comparison metrics, such as the Davies-Bouldin and Calinski-Harabasz indices, with the advantage of providing built-in normalization that makes it ideal for comparing between unique model evaluations.[19] This process determined an optimized unsupervised clustering method, which uses the Bray–Curtis distance metric for dimensionality reduction with UMAP and gaussian mixture model clustering (optimal parameters in

Supplemental Table S4). The optimal model had four distinct clusters, Supplemental Figure S1, which were used for phenotype and outcome assessment in the training and testing populations. The image-only cluster model identified two clusters as the optimal number, but with lower Silhouette score (0.670).

To validate the outcome assessment, the unsupervised clustering model was deployed in the external cohort. To infer cluster assignments during validation, the trained dimensionality reduction model maps individual patients into the embedding space, and cluster assignments are provided based on the nearest cluster centroids recorded during training. A hierarchical clustering algorithm was applied to the cluster assignments to generate a dendrogram based on distances in Euclidean space.

Stability analysis

The overall stability of the cluster model was assessed by evaluating the classification of patients across bootstrapped patient groups from the training population. Bootstrapping was performed 100 times with 100 patients from the external testing population in each sample to strenuously evaluate the reclassification of samples from the external testing population. Jaccard

index and adjusted Rand index were used to quantify multiclass reclassification accuracy.[20,21]

Model explainability

In order to address the black box nature of the unsupervised cluster analysis, we applied Shapley additive explanations (SHAP) to determine which variables contributed most to the cluster assignments.[22] In this analysis, we determined SHAP values (which reflect the relative contribution of a variable to the overall model) for the cluster model developed in the training population.

Statistics

Normally distributed data are presented with mean and standard deviation. Data that are not normally distributed are presented as median and interquartile range (IQR). Categorical data are presented as number and percentage. Statistical significance was assessed using Mann–Whitney Wilcoxon, Kruskal–Wallis rank sum, or Pearson's chi-squared test. The study employed an opportunistic sample size since the data was already available. We included all eligible patients from the REFINE SPECT registry to ensure a large, representative population was available for model training and testing. The hypergeometric distribution v-test (R, Version 4.1.1; FactoMineR package, Version 2.4) was performed to assess the representation of variables within each cluster, with a positive v-test score indicating over-representation of the variable within the cluster and a negative v-test score indicating under-representation.[23]

Cox proportional hazards analysis was used to assess associations with death or non-fatal myocardial infarction using hazard ratios (HR) and 95% confidence intervals (CI). We compared the risk associated with clusters to traditional risk factors, including diabetes, previous MI, and pharmacologic stress. We also evaluated for possible difference by sex. Additionally, we evaluated the independent prognostic significance of the Clusters after adjusting for typical clinical variables, including age, sex, medical history (hypertension, diabetes, dyslipidemia, known coronary disease), and stress perfusion abnormality. We also evaluated the independent prognostic value compared to an extended multivariable model in the external testing population. Variables were selected based on known clinical importance, with no dedicated model building performed. The two models were evaluated to assess the independent importance of the Clusters, which would allow for appropriate adjustment while minimizing multicollinearity, since variables are incorporated into the clusters. In the multivariable models, we evaluated differences in the associations between cluster and outcomes, as a function of site, using interaction terms. The benefit of including the cluster model with the extended multivariable model was evaluated with improvement in model fit using likelihood ratio chi-square and patient classification using continuous net-reclassification index. Kaplan–Meier curves were used to visualize rates of the composite outcome. A two-sided p value < 0.05 was considered statistically significant. Analyses were performed using Stata version 14.2 (StataCorp, College Station, Texas) and R 4.1.1.

Ethics approval and reporting guidelines

This study complied with the Declaration of Helsinki and was approved by the institutional review boards at each participating institution, with the overall study approved at Cedars-Sinai Medical Center (REB ID # Pro00019604). This study was designed and conducted following the Proposed Requirements for Cardiovascular Imaging–Related Machine Learning Evaluation,[24] with the checklist included as Supplemental Table S5.

Role of funders

Results

Patient population

A total of 9849 patients were included in the training population and 12,528 patients in the external testing population. An overview of the study findings is shown in Fig. 2. Population characteristics are outlined in Table 1. There were statistically significant differences in characteristics between the external testing and training populations, including older age (median 65 vs 63, p < 0.001) and a higher proportion of patients undergoing pharmacologic stress (48.5% vs 43.6%, p < 0.001) in the external testing population. Overall, 0.55% of values were missing in the training population and 1.51% in the external population, with details in Supplemental Table S6.

Training population

The algorithm identified four distinct clusters. These four clusters were used for phenotype and outcome assessment in the training and testing populations. Characteristics of the clusters in the training population are outlined in Table 2 and the embedding space components are visualized in Fig. 3. Patients in Clusters 1
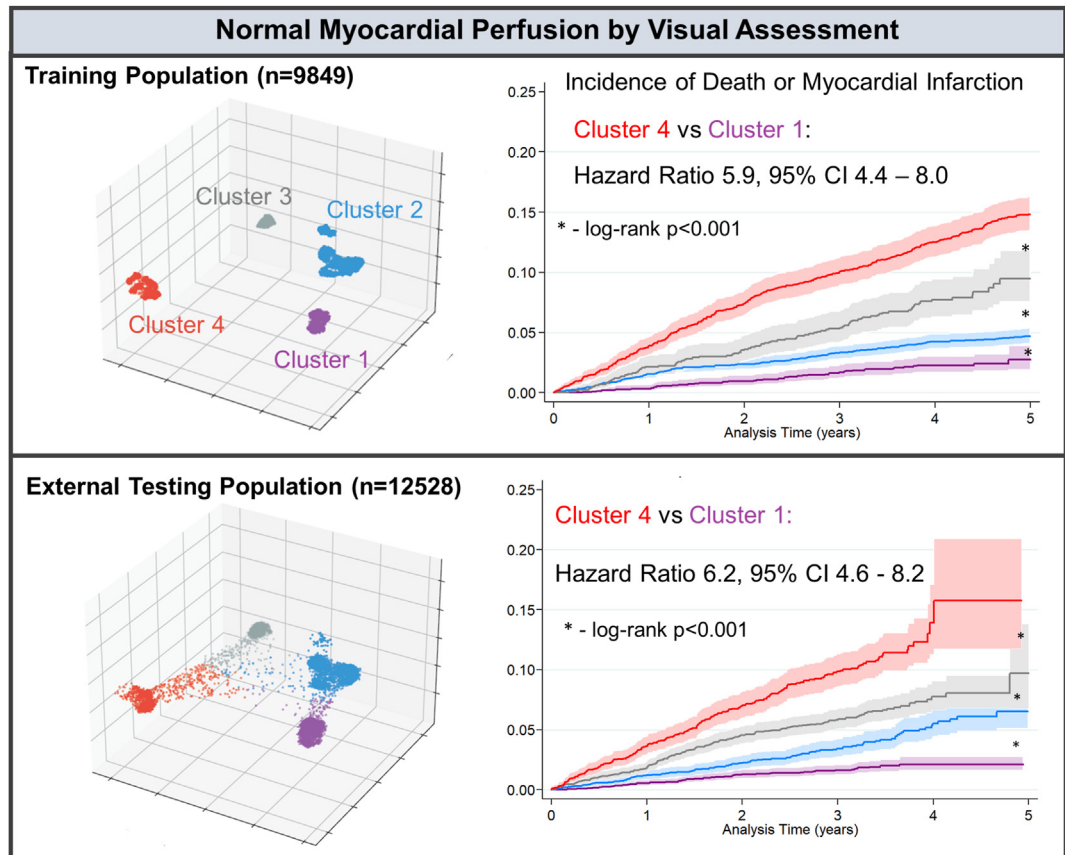
Fig. 2: Overview of study design and findings. Unsupervised machine learning was used to identify four unique clusters of patients with normal myocardial perfusion. Patients in the highest risk cluster had 6-fold higher risk of death or myocardial infarction in the training and external testing populations.

and 2 were more likely to have completed exercise stress (89%) compared to patients in Clusters 3 or 4 (0.0%, p < 0.001). Patients in Cluster 2 were more likely to have an ischemic clinical response to stress (chest pain) (42.0 vs 0.1%, p < 0.001) and stress perfusion abnormality >5% (7.7% vs 3.3%, p < 0.001) compared to Cluster 1. However, phase bandwidth was lower in Cluster 2 compared to Cluster 1 (median 24 vs 30, p < 0.001). Patients in Cluster 4 were more likely to have stress perfusion abnormality >5% (15.0% vs 4.4%, p < 0.001), left ventricular dilation (14.7% vs 10.3%, p < 0.001), and left ventricular ejection fraction <50% (6.4% vs 1.5%, p < 0.001) compared to Cluster 3. The model's silhouette score was 0.881 (Supplemental Figure S1). Further differences in quantitative perfusion parameters between clusters are visualized in Supplemental Figure S2A–L. Shapley additive explanation values for variables in the training population are shown in Fig. 4.

During a median follow-up of 5.3 (IQR 4.2–6.5) years, 801 (8.1%) patients experienced at least one event, including a total of 715 deaths (7.3%) and 106 non-fatal

myocardial infarctions (1.1%). Kaplan–Meier curves for death or myocardial infarction in the training population, stratified by cluster, are shown in Fig. 5. Patients in Cluster 4 were at the highest risk (HR 5.94, 95% CI 4.40–8.03) followed by Cluster 3 (HR 3.53, 95% CI 2.49–5.00), Cluster 2 (HR 1.82, 95% CI 1.33–2.49), and Cluster 1 (reference group, p < 0.001 for all comparisons). The risk associated with Cluster 4 was similar in female (HR 6.80, 95% CI 4.19–11.0) and male patients (HR 5.72, 95% CI 3.88–8.42; interaction p-value 0.503). In the multivariable model (Supplemental Table S7), patients in Cluster 4 were still at statistically significantly higher risk than patients in Cluster 1 (adjusted HR 4.37, 95% CI 3.21–5.94, p < 0.001).

External testing population
Characteristics of the four clusters in the external testing population are outlined in Table 3, and the embedding space components are visualized in Fig. 3. The silhouette width in the external population was 0.852, suggesting that patients continued to be categorized into

	Training population N = 9849	External testing population N = 12,528
Age	63 (54–71)	65 (57–73)
Male	4657 (47.3%)	6141 (49.0%)
Body mass index	28.1 (25.0, 32.2)	27.9 (24.8, 32.0)
Past medical history		
Past myocardial infarction	539 (5.5%)	839 (6.7%)
Past PCI or Stents	1032 (10.5%)	999 (8.0%)
Past CABG	390 (4.0%)	341 (2.7%)
Hypertension	6087 (61.8%)	7657 (61.1%)
Diabetes Mellitus	2326 (23.6%)	2886 (23.0%)
Dyslipidemia	5868 (59.6%)	4952 (39.5%)
Family History	2868 (29.1%)	5100 (40.7%)
Presenting symptoms		
Typical chest pain	509 (5.1%)	369 (3.1%)
Atypical chest pain	2314 (23.5%)	5137 (41.0%)
Location		
Emergency	394 (4.0%)	509 (4.4%)
Inpatient	884 (9.0%)	1305 (11.4%)
Outpatient	8570 (87%)	9648 (84.2%)
Mode of stress		
Exercise	5558 (56.4%)	6449 (51.5%)
Pharmacologic	4291 (43.6%)	6077 (48.5%)
Clinical response to stress		
Abnormal	388 (4.0%)	314 (2.5%)
Equivocal/Non-diagnostic	1427 (14.8%)	862 (6.9%)
Ischemic	2384 (24.2%)	418 (3.3%)
Non-Ischemic	5605 (57.0%)	10,934 (87.3%)

Continuous variables presented as median (interquartile range) and categorical variables as number (proportion). CABG—coronary artery bypass grafting, ECG—electrocardiogram, PCI—percutaneous coronary intervention.

Table 1: **Population characteristics for the training and external testing populations.**

distinct classes using the previously trained model. Patients in clusters 1 and 2 were more likely to have completed exercise stress (96.2%, p < 0.001) compared to patients in clusters 3 or 4 (0.2%, p < 0.001). Patients in Cluster 2 were more likely to have left ventricular ejection fraction <50% (7.6% vs 3.7%, p < 0.001) compared to Cluster 1. Patients in Cluster 3 were more likely to present with typical (5.7% vs 0.1%, p < 0.001) or atypical chest pain (53.8% vs 36.6%, p < 0.001). Patients in Cluster 4 had higher median BMI compared to Cluster 3 (30 vs 27, p < 0.001), were more likely to have left ventricular dilation (19.3% vs 11.1%, p < 0.001), and left ventricular ejection fraction <50% (7.2% vs 5.9%, p = 0.047). Further differences in quantitative perfusion parameters between clusters are shown in Supplemental Figure 3A–L.

During median follow-up of 3.1 (IQR 2.3–3.6) years, 590 (4.7%) patients experienced at least one event, including a total of 438 deaths (3.5%) and 173 non-fatal myocardial infarctions (1.4%). Kaplan–Meier curves for death or non-fatal myocardial infarction in the testing population are shown in Fig. 6. Patients in Cluster 4 were at the highest risk (HR 6.17, 95% CI 4.64–8.20) followed by Cluster 3 (HR 3.56, 95% CI 2.67–4.74), Cluster 2 (HR 2.18, 95% CI 1.59–2.99), and Cluster 1 (reference group, p < 0.001 for all comparisons). The risk associated with Cluster 4 was similar in female (HR 6.51, 95% CI 4.12–10.3) and male patients (HR 6.78, 95% CI 4.70–9.79; interaction p-value 0.906). For comparison, the risk associated with diabetes (HR 1.68, 95% CI 1.42–2.00), pharmacologic stress (HR 3.03, 95% CI 2.53–3.63), and previous MI (HR 1.82, 95% CI 1.40–2.36) were all lower than the risk associated with Cluster 4 vs Cluster 1.

In the multivariable analysis, patients in Cluster 4 were at statistically significantly higher risk than patients in Cluster 1 (adjusted HR 5.23, 95% CI 3.91–6.99, p < 0.001). Clusters 2 and 4 were associated with an increased risk of adverse events in a more extensive multivariable model (Supplemental Table S8). Including the clusters also improved model fit (increase in LR chi-square 68.9, p < 0.001) and patient classification (continuous net reclassification index 0.213, 95% CI 0.131–0.295).

Stability analysis

Supplemental Table S3 shows the result of clustering embeddings from the optimal, trained UMAP model for the three clustering algorithms. We observed similar silhouette scores from all three clustering algorithms for the internal cohort at four clusters, highlighting the stability of the selected model.

The Jaccard index and adjusted Rand index demonstrated good stability with respect to patient classification with values of 0.993 ± 0.011 and 0.991 ± 0.016, respectively, where a value of 1 represents perfect agreement in patient classification across all bootstrapped samples. We also evaluated the absolute distance from assigned clusters using Euclidean distance, as shown in Supplemental Figure S4. There was a similar proportion of patients classified as outliers (defined as Euclidean distance >3 standard deviations above the mean) in the internal (6.8%) and external populations (5.9%). A dendrogram is presented in Supplemental Figure S5. Results for the imaging-only analysis are available in the Supplemental Results and Supplemental Figure S6.

Discussion

Leveraging unsupervised machine learning techniques, we identified four distinct phenotypes among patients with normal myocardial perfusion, with clear differences between groups in both internal and external populations. While many characteristics differed, the use of exercise vs pharmacologic stress was the most important variable for splitting the low-risk from high-risk groups. Importantly, despite

|	Cluster 1 N = 2097	Cluster 2 N = 4148	Cluster 3 N = 1165	Cluster 4 N = 2439
Age	61 (53, 67)	60 (51, 68)	69 (63, 77)	65 (56, 74)
Male	1072 (51%)	2167 (52%)	421 (36%)	997 (41%)
Body mass index	27 (25, 29)	28 (25, 33)	28 (25, 31)	30 (25, 36)
Past medical history				
Past myocardial infarction	70 (3.3%)	215 (5.2%)	56 (4.8%)	198 (8.1%)
Past PCI or Stents	215 (10%)	384 (9.3%)	177 (15%)	256 (10%)
Past CABG	49 (2.3%)	167 (4.0%)	58 (5.0%)	116 (4.8%)
Hypertension	1003 (48%)	2419 (58%)	841 (72%)	1824 (75%)
Diabetes Mellitus	419 (20%)	779 (19%)	331 (28%)	797 (33%)
Dyslipidemia	1285 (61%)	2370 (57%)	793 (68%)	1420 (58%)
Family History	635 (30%)	1461 (35%)	203 (17%)	569 (23%)
Presenting symptoms				
Typical chest pain	25 (1.2%)	338 (8.2%)	24 (2.1%)	122 (5.0%)
Atypical chest pain	240 (11.4%)	1160 (28.0%)	143 (12.3%)	771 (31.6%)
Location				
Emergency	0 (0%)	297 (7.2%)	0 (0%)	97 (4.0%)
Inpatient	0 (0%)	381 (9.2%)	0 (0%)	503 (21%)
Outpatient	2097 (100%)	3470 (84%)	1165 (100%)	1838 (75%)
Mode of stress				
Exercise	2097 (100%)	3461 (83%)	0 (0%)	0 (0%)
Pharmacologic	0 (0%)	687 (17%)	1165 (100%)	2437 (100%)
Clinical response to stress				
Abnormal	43 (2.2%)	298 (7.4%)	42 (3.9%)	5 (0.3%)
Equivocal/Non-Diagnostic	1 (<0.1%)	208 (5.1%)	0 (0%)	314 (17%)
Ischemic	2 (0.1%)	1700 (42%)	6 (0.6%)	676 (38%)
Non-Ischemic	1915 (98%)	1844 (46%)	1041 (96%)	805 (45%)
Outcomes				
Death	32 (1.5%)	220 (5.3%)	76 (6.5%)	387 (16%)
Myocardial Infarction	16 (0.8%)	29 (0.7%)	17 (1.5%)	44 (1.8%)

Continuous variables presented as median (interquartile range) and categorical variables as number (proportion). CABG—coronary artery bypass grafting, ECG—electrocardiogram, PCI—percutaneous coronary intervention.

Table 2: **Patient characteristics in each of the clusters in the training population.**

having normal myocardial perfusion, patients in Cluster 4 had a 6-fold higher risk of death or non-fatal myocardial infarction, compared to patients in Cluster 1. The risk associated with this large cluster (>20% of patients) was substantially higher compared to traditional clinical subgroups such as history of diabetes or previous myocardial infarction. Such artificial intelligence-based clustering could be used clinically to improve classification of ~90% of the 15–20 million scans performed annually.[1] Physicians could be presented with a patient's cluster assignment in conjunction with the MPI report, potentially allowing them to target aggressive therapies more accurately in the highest risk groups and provide additional reassurance to the lowest risk groups.

The unsupervised model identified 4 patient clusters, which provided optimal differentiation between groups. Importantly, our study is one of very few to demonstrate that these differences are maintained when the trained model is applied to an external testing population. Two

of the clusters (Clusters 1 and 2) were primarily made up of patients who were able to complete exercise stress, while the remaining two (Clusters 3 and 4) comprised patients who almost exclusively required pharmacologic stress. Similarly, 13 of the top 24 features determining cluster assignment were related to the mode of stress. This is likely because other features, such as peak stress heart rate and blood pressure, are directly related to the mode of stress. Regardless, this finding confirms the known importance of a patient's ability to complete exercise stress.[26] Aside from stress modality, there were smaller differences in traditionally important clinical parameters such as diabetes[8] or previous myocardial infarction.[9] In the external testing set, patients in Cluster 3 were more likely to be presenting with typical or atypical chest pain. We also noted statistically significant differences in markers of ventricular morphology between clusters, such as higher stress shape index in Cluster 4 and Cluster 2 compared to Cluster 3 and Cluster 1. The differences in shape index could help
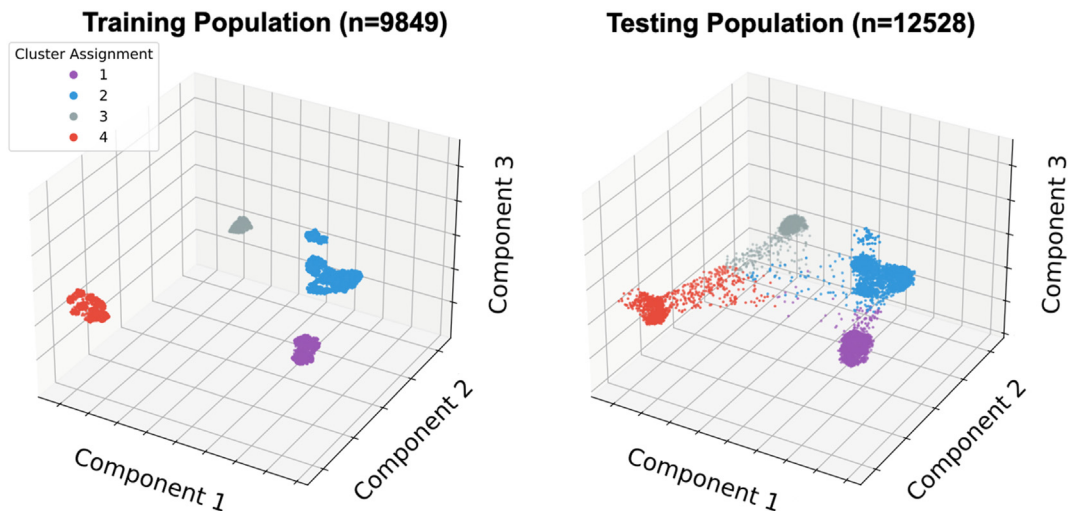
Fig. 3: Training and external testing populations projected into the reduced embedding space. Components of the embedding space are independent summary measures that combine multiple input parameters, which are determined by the non-linear dimensionality reduction process. There is a clear separation of groups in the training population with Silhouette score of 0.881. In the external testing population, there is still separation between groups, but with a few outliers, the Silhouette score was 0.852.

explain differences in risk between clusters.[16] However, the variation seen in other image parameters like phase bandwidth differed in a way that would be expected to decrease the risk associated with the highest risk cluster.[15] There were minor discrepancies in stress perfusion abnormalities, with differences potentially minimized by only including patients with visually normal perfusion. Additionally, clusters were less distinct when we applied the same process to imaging variables alone. Such complex variations of multiple clinical and imaging variables could not be easily captured with standard clinical rules. It may be more practical for clinical use to reference risk phenotypes, in this case a patient's cluster assignment, instead of listing multiple individual features for risk predictions. Overall, these findings highlight the potential value of comprehensive patient phenotyping, using unsupervised learning, for patient classification and risk prediction.

Unsupervised learning represents a potentially valuable artificial intelligence technique which has not been frequently applied to cardiac imaging.[27–29] Unsupervised algorithms learn and label inherent relationships or structure within data,[12] allowing identification of groups without preset expectations. A key potential benefit of using unsupervised learning to phenotype patients is that patients are characterized using features readily available at the time of study reporting. There have been a few recent applications of unsupervised approaches in cardiac imaging. Lancaster et al. identified clusters of patients with a higher risk of death among 866 patients undergoing echocardiography.[30] Krohn et al. applied unsupervised learning to

identify phenotypes among patients undergoing invasive coronary angiography.[31] Shah et al. applied hierarchical cluster analysis to identify 3 clusters among 397 patients with heart failure preserved ejection fraction undergoing echocardiography.[32] The authors went on to demonstrate that the risk of heart failure hospitalization associated with different clusters in a separate internal testing group (n = 107).[32] Importantly, we demonstrate unsupervised learning in a substantially larger number of patients from an international registry including 10 sites and reproduced the results in an external (previously unseen by the trained model) population from independent centers. External testing is critical for unsupervised learning because some imaging data is related to site-specific protocols for training; for example, stress dose is related to the use of stress-first or stress-only imaging. The fact that clustering maintains performance in a large external and multi-center testing cohort suggests that our approach could be applied to most (over 90% in some cohorts[5]) of the 15–20 million cardiac perfusion scans performed annually worldwide,[1] of which a vast proportion have normal results.

Our approach can help physicians identify important subgroups within the population of patients with normal myocardial perfusion. This is a critical population to study because these patients are often assigned the same risk estimate unless other markers of increased risk are present. Importantly, outcomes have not been used at any point during data training, unlike in traditional machine learning or statistical analysis with survival models. It is possible that patients in the clusters have distinct underlying disease processes or
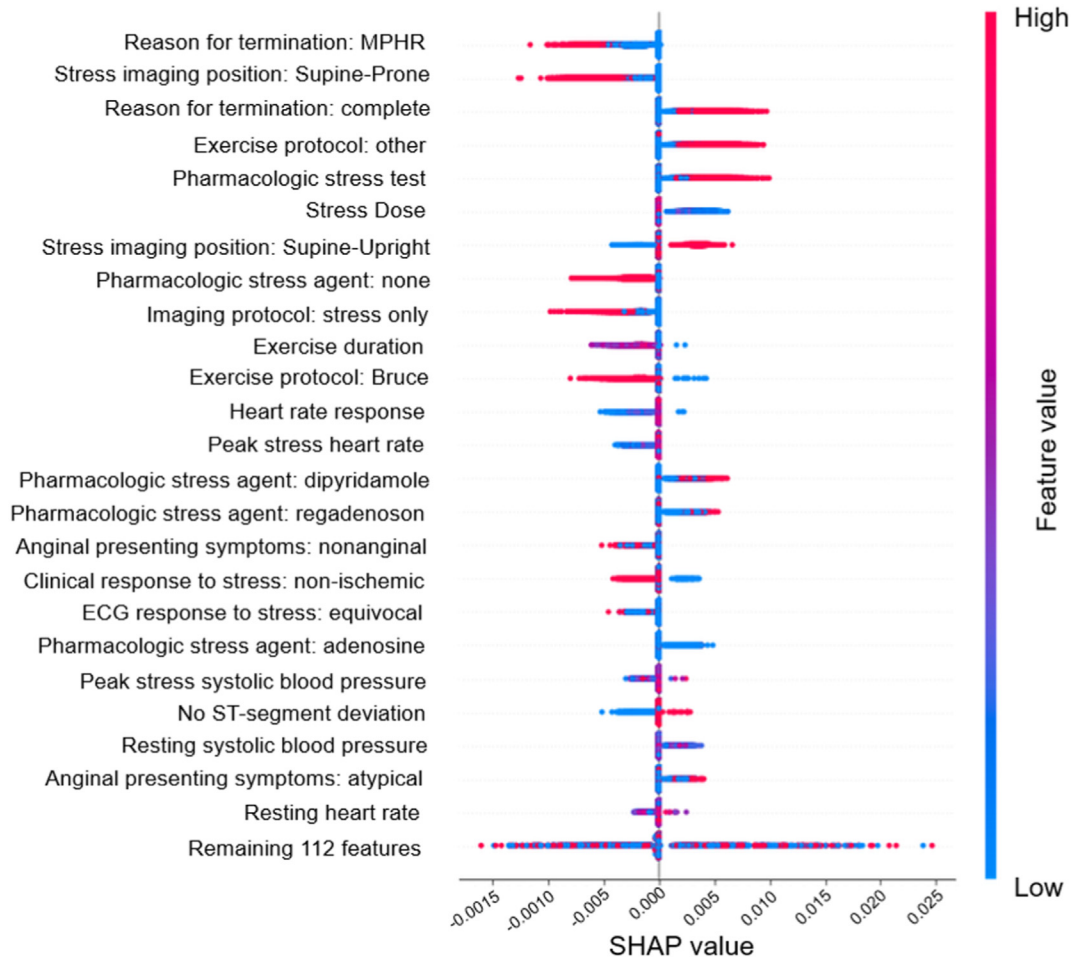
Fig. 4: Shapley additive explanation (SHAP) values for the top 24 features and all remaining features were generated using the Cluster-Shapley method.[25] ECG—electrocardiogram, MPHR—maximum predicted heart rate.

other unmeasured variables that may be leading to the classifications. For example, frailty may be an important confounding factor, which plays an important role in leading patients to require pharmacologic stress and would be associated with cardiovascular events. The prevalence of non-perfusion high-risk markers, such as impaired left ventricular systolic function,[10] is relatively low in this patient population. Additionally, the significance of findings such as transient ischemic dilation has been called into question in this population.[33] Traditionally, physicians have relied on other clinical features, such as medical history, to identify potentially higher risk patients; however, the associated risk for Cluster 4 in our study was higher than the risk associated with features such as requirement for pharmacologic stress or diabetes. The unsupervised approach is efficiently combining some of these important features, removing the need for clinicians to attempt to integrate all this information on their own.

Study limitations

Our study is not without its limitations. We do not have information on how physicians responded to MPI results and changes in therapy in response to clinical or imaging features. If physicians reacted to specific features this would bias our estimate of associations with outcomes. Unsupervised learning is specifically designed to identify distinct patient phenotypes and is not optimized to provide stratification for a specific outcome as they are not used for model training. However, we included this analysis to demonstrate the potential clinical application of this approach. In comparison, supervised approaches depend on the availability and quality of outcome data in the training set but do not establish patient phenotypes. We utilized UMAP for dimensionality reduction, and it is possible that different optimal clusters would be identified if another reduction approach was used; however, we verified that the number of clusters was similar across downstream
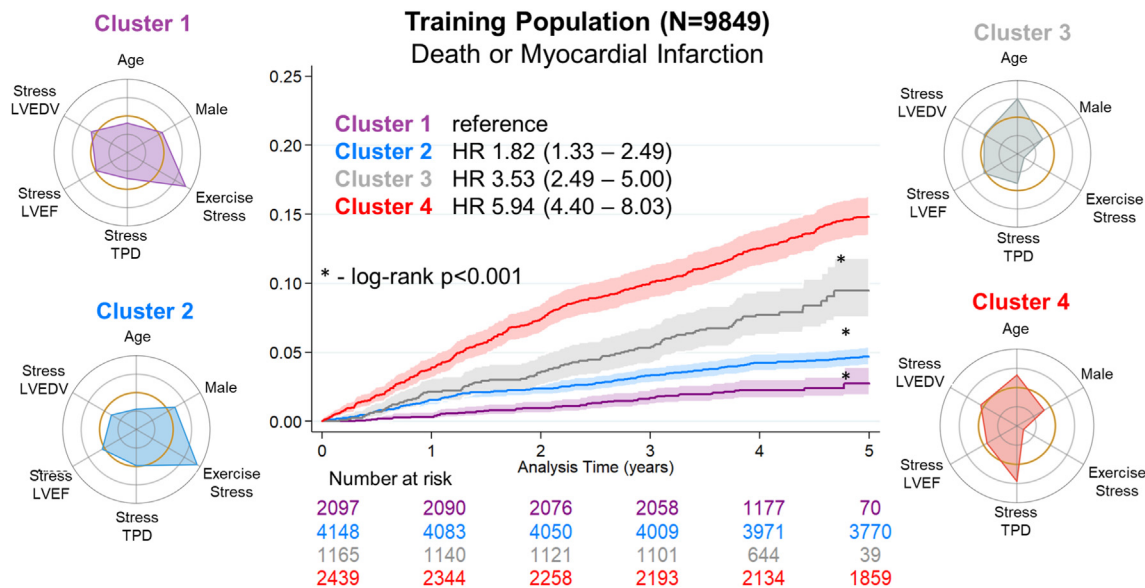
Fig. 5: Kaplan–Meier curves for death or myocardial infarction in the training population. The radial plots visualize differences between clusters compared to the entire training population (inside/outside of orange circle). A higher proportion of patients, or larger values, is identified as an over-represented trait (outside). The opposite is true for an under-represented trait (inside). Hazard ratio (HR) and 95% confidence intervals are shown for each Cluster compared to Cluster 1. LVEDV—left ventricular end-diastolic volume, LVEF—left ventricular ejection fraction, TPD—total perfusion deficit.

clustering methods. UMAP performance is dependent on hyperparameter optimization according to the target dataset. In this study, we carried out an extensive grid search to optimize model performance for our training data, and we recommend a similarly robust search be carried out if extending this approach to new applications. We utilized Shapley additive explanation values to identify the most important features leading to cluster assignments, which could also be utilized for patient-specific explanations. We do not have information regarding serial changes in perfusion or ventricular function, which may be relevant for identifying important subgroups of patients.[34] We included patients from separate centers and some clinical features, such as the use of exercise vs pharmacologic stress, vary between sites. This is particularly evident when comparing the training and external testing populations, where statistically significant differences were present in almost all clinical features. Indeed, this variation in population characteristics is critical for external validation to ensure results are broadly generalizable and the model can be applied to unseen centers. While the overall proportion of missing values was small, imputation of missing values may have decreased the differences between groups. Lastly, there was a difference in follow-up duration between sites, as evidenced by changes in patients at risk over time on the Kaplan–Meier curves. However, this heterogeneity increases the generalizability of our findings, as demonstrated by the robust results in external testing, which is a critical step for any

model.[35] Lastly, we used all-cause mortality as part of the outcome, and results may have been different if cardiovascular mortality was used.

Conclusion

Unsupervised learning identified four patient phenotypes among patients with normal myocardial perfusion. Using a large, multicenter, international external testing population, these clusters identified patients with substantially higher risk of death or myocardial infarction compared to traditional predictors. Our results suggest a potential role for patient phenotyping to improve risk stratification of patients with normal imaging results.

	Cluster 1 N = 3457	Cluster 2 N = 3237	Cluster 3 N = 3299	Cluster 4 N = 2535
Age	63 (55, 70)	61 (53, 69)	69 (61, 76)	68 (59, 75)
Male	1850 (54%)	1767 (55%)	1546 (47%)	978 (39%)
Body mass index	28 (25, 31)	28 (25, 32)	27 (24, 30)	30 (26, 35)
Past medical history				
Past myocardial infarction	200 (5.8%)	159 (4.9%)	385 (12%)	95 (3.7%)
Past PCI or Stents	233 (6.7%)	230 (7.1%)	377 (11%)	159 (6.3%)
Past CABG	86 (2.5%)	70 (2.2%)	112 (3.4%)	73 (2.9%)
Hypertension	1874 (54%)	1677 (52%)	2353 (71%)	1753 (69%)
Diabetes Mellitus	703 (20%)	561 (17%)	823 (25%)	816 (32%)
Dyslipidemia	827 (24%)	1403 (43%)	1728 (52%)	994 (39%)
Family History	1811 (52%)	1316 (41%)	1340 (41%)	633 (25%)
Smoking	1290 (37%)	595 (18%)	1438 (44%)	661 (26%)
Presenting symptoms				
Typical chest pain	91 (3.1%)	97 (3.0%)	179 (5.7%)	2 (0.1%)
Atypical chest pain	1175 (39.5%)	1389 (43.2%)	1680 (53.8%)	893 (36.6%)
Location				
Emergency	30 (0.9%)	220 (12%)	18 (0.6%)	241 (13%)
Inpatient	199 (5.9%)	101 (5.5%)	429 (15%)	135 (7.1%)
Outpatient	3159 (93%)	1500 (82%)	2461 (85%)	1516 (80%)
Mode of Stress				
Exercise	3389 (98%)	3050 (94%)	5 (0.2%)	5 (0.2%)
Pharmacologic	68 (2.0%)	185 (5.7%)	3294 (>99%)	2530 (>99%)
Clinical response to stress				
Abnormal	75 (2.2%)	196 (6.3%)	3 (<0.1%)	40 (1.7%)
Equivocal/Non-Diagnostic	34 (1.0%)	241 (7.8%)	36 (1.1%)	103 (4.4%)
Ischemic	226 (6.7%)	110 (3.6%)	20 (0.6%)	62 (2.7%)
Non-Ischemic	3044 (90%)	2543 (82%)	3234 (98%)	2113 (91%)
Outcomes				
Death	35 (1.0%)	81 (2.5%)	164 (5.0%)	158 (6.2%)
Myocardial Infarction	27 (1.9%)	31 (1.0%)	46 (1.6%)	69 (3.9%)

Continuous variables presented as median (interquartile range) and categorical variables as number (proportion). CABG—coronary artery bypass grafting, ECG—electrocardiogram, PCI—percutaneous coronary intervention.

Table 3: **Patient characteristics in each of the clusters in the external testing population.**

Fig. 6: Kaplan–Meier curves for death or myocardial infarction in the external testing population. The radial plots visualize differences between clusters compared to the entire external testing population (inside/outside of orange circle). A higher proportion of patients, or larger values, is identified as an over-represented trait (outside). The opposite is true for an under-represented trait (inside). Hazard ratio (HR) and 95% confidence intervals are shown for each Cluster compared to Cluster 1. LVEDV—left ventricular end-diastolic volume, LVEF—left ventricular ejection fraction, TPD—total perfusion deficit.

References
1 Einstein AJ. Multiple opportunities to reduce radiation dose from myocardial perfusion imaging. *Eur J Nucl Med Mol Imaging.* 2013;40(5):649–651.
2 Gulati M, Levy PD, Mukherjee D, et al. 2021 AHA/ACC/ASE/CHEST/SAEM/SCCT/SCMR guideline for the evaluation and diagnosis of chest pain. *Circulation.* 2021;144(22):e368–e454.
3 Otaki Y, Betancur J, Sharir T, et al. 5-Year prognostic value of quantitative versus visual MPI in subtle perfusion defects. *JACC Cardiovasc Imaging.* 2020;13(3):774–785.
4 Berman DS, Abidov A, Kang X, et al. Prognostic validation of a 17-segment score derived from a 20-segment score for myocardial perfusion SPECT interpretation. *J Nucl Cardiol.* 2004;11(4):414–423.
5 Rozanski A, Gransar H, Hayes SW, et al. Temporal trends in the frequency of inducible myocardial ischemia during cardiac stress testing: 1991 to 2009. *J Am Coll Cardiol.* 2013;61(10):1054–1065.
6 Al Badarin FJ, Chan PS, Spertus JA, et al. Temporal trends in test utilization and prevalence of ischaemia with positron emission tomography myocardial perfusion imaging. *Eur Heart J Cardiovasc Imaging.* 2020;21(3):318–325.
7 Einstein AJ, Pascual TN, Mercuri M, et al. Current worldwide nuclear cardiology practices and radiation exposure. *Eur Heart J.* 2015;36(26):1689–1696.
8 Han D, Rozanski A, Gransar H, et al. Myocardial ischemic burden and differences in prognosis among patients with and without diabetes. *Diab Care.* 2020;43(2):453–459.
9 Miller RJH, Klein E, Gransar H, et al. Prognostic significance of previous myocardial infarction and previous revascularization in patients undergoing SPECT MPI. *Int J Cardiol.* 2020;313:9–15.
10 Sharir T, Germano G, Kavanagh PB, et al. Incremental prognostic value of post-stress left ventricular ejection fraction and volume by gated myocardial perfusion single photon emission computed tomography. *Circulation.* 1999;100(10):1035–1042.
11 Miller RJH, Hu LH, Gransar H, et al. Transient ischaemic dilation and post-stress wall motion abnormality increase risk in patients with less than moderate ischaemia: analysis of the REFINE SPECT registry. *Eur Heart J Cardiovasc Imaging.* 2020;21(5):567–575.
12 Miller RJH, Huang C, Liang JX, Slomka PJ. Artificial intelligence for disease diagnosis and risk prediction in nuclear cardiology. *J Nucl Cardiol.* 2022;29(4):1754–1762.
13 Segar MW, Patel KV, Ayers C, et al. Phenomapping of patients with heart failure with preserved ejection fraction using machine learning-based unsupervised cluster analysis. *Eur J Heart Fail.* 2020;22(1):148–158.
14 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.* 2020;27(3):1010–1021.
15 Kuronuma K, Miller RJH, Otaki Y, et al. Prognostic value of phase analysis for predicting adverse cardiac events beyond conventional single-photon emission computed tomography variables. *Circ Cardiovasc Imaging.* 2021;14(7):e012386.
16 Miller RJH, Sharir T, Otaki Y, et al. Quantitation of poststress change in ventricular morphology improves risk stratification. *J Nucl Med.* 2021;62(11):1582–1590.
17 McInnes L, Healy J, Melville J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv.* 2018. https://doi.org/10.48550/arXiv.1802.03426.

18 Cao J, Spielmann M, Qiu X, et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature*. 2019;566(7745):496–502.

19 Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math*. 1987;20:53–65.

20 Jaccard P. The distribution of the flora in the Alpine Zone. *New Phytol*. 1912;11(2):37–50.

21 Rand WM. Objective criteria for the evaluation of clustering methods. *J Am Stats Assoc*. 1971;66(336):846–850.

22 Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Advanc Neur Inf Process*. 2017;30.

23 Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. *J Stat Software*. 2008;25(1):1–18.

24 Sengupta PP, Shrestha S, Berthon B, et al. Proposed requirements for cardiovascular imaging-related machine learning evaluation (PRIME). *JACC Cardiovasc Imaging*. 2020;13(9):2017–2035.

25 Marcilio-Jr WE, Eler DM. Explaining dimensionality reduction results using Shapley values. *Expert System Application*. 2021;178: 115020.

26 Rozanski A, Gransar H, Hayes SW, Friedman JD, Hachamovitch R, Berman DS. Comparison of long-term mortality risk following normal exercise vs adenosine myocardial perfusion SPECT. *J Nucl Cardiol*. 2010;17(6):999–1008.

27 Cho Jung S, Shrestha S, Kagiyama N, et al. A network-based "phenomics" approach for discovering patient subtypes from high-throughput cardiac imaging data. *JACC Cardiovasc Imaging*. 2020;13(8):1655–1670.

28 Pezel T, Unterseeh T, Hovasse T, et al. Phenotypic clustering of patients with newly diagnosed coronary artery disease using cardiovascular magnetic resonance and coronary computed tomography angiography. *Front Cardiovasc Med*. 2021;8:760120.

29 Yoon YE, Baskaran L, Lee BC, et al. Differential progression of coronary atherosclerosis according to plaque composition: a cluster analysis of PARADIGM registry data. *Sci Rep*. 2021;11(1):17121.

30 Lancaster MC, Salem Omar AM, Narula S, Kulkarni H, Narula J, Sengupta PP. Phenotypic clustering of left ventricular diastolic function parameters: patterns and prognostic relevance. *JACC Cardiovasc Imaging*. 2019;12(7, Part 1):1149–1161.

31 Krohn JB, Nguyen YN, Akhavanpoor M, et al. Identification of specific coronary artery disease phenotypes implicating differential pathophysiologies. *Front Cardiovasc Med*. 2022;9:778206.

32 Shah SJ, Katz DH, Selvaraj S, et al. Phenomapping for novel classification of heart failure with preserved ejection fraction. *Circulation*. 2015;131(3):269–279.

33 Valdiviezo C, Motivala AA, Hachamovitch R, et al. The significance of transient ischemic dilation in the setting of otherwise normal SPECT radionuclide myocardial perfusion images. *J Nucl Cardiol*. 2011;18(2):220–229.

34 Miller RJH, Nabipoor M, Youngson E, et al. Heart failure with mildly reduced ejection fraction: retrospective study of ejection fraction trajectory risk. *ESC Heart Fail*. 2022;9(3):1564–1573.

35 Siontis GC, Tzoulaki I, Castaldi PJ, Ioannidis JP. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol*. 2015;68(1):25–34.