

AI Characterisation of Discordance Profiles Between Stress Electrocardiogram and Myocardial Tomoscintigraphy Using Random Forest XGBoost and SHAP

Youness El Maadaoui¹, Abdelaziz Belaguid², Mohamed Aziz Bsiss³, Aboubaker Matrane³

¹Electronic Systems Sensors and Nanobiotechnology, National School of Arts and Crafts, Mohammed v University, Rabat, Morocco; ²Department of Physiology, Faculty of Medicine and Pharmacy, Rabat, Morocco; ³Department of Nuclear Medicine, MOHAMMED VI University Hospital, Cadi Ayyad University, Marrakesh, Morocco

Correspondence: Youness El Maadaoui, National School of Arts and Crafts, Mohammed v University in Rabat, Avenue de l'Armée Royale, Madinat Al Irfane BP 6207, Rabat, 10000, Morocco, Email youness_elmaadaoui@um5.ac.ma

Purpose: The evaluation of Coronary Artery Disease (CAD) using stress ECG and Myocardial Perfusion SPECT (MPS) frequently reveals discrepancies, particularly in patients with a positive ECG and a negative MPS (MPS-/ECG+). This paradoxical group poses a diagnostic challenge, often unexplained by classic statistical analyses. Our study aims to characterize this profile using interpretable Artificial Intelligence (AI).

Patients and Methods: A cohort of 86 patients with negative MPS was stratified into a “discordant” group (MPS-/ECG+, n=19) and a “concordant” group (MPS-/ECG-, n=67). A two-pronged analytical approach was used: (1) bivariate statistical analysis and (2) machine learning modelling. Two algorithms, Random Forest and XGBoost, were trained on a dataset rebalanced using SMOTE. Performance was assessed by AUC, and interpretability was ensured by SHAP analysis.

Results: Conventional univariate statistical analysis did not identify any variable significantly associated with discordance (all $p > 0.05$). In contrast, the XGBoost model, exploiting multivariate interactions, surpassed Random Forest, achieving a moderate but informative performance (AUC = 0.78), with a Sensitivity (Recall) of 67%, a Precision (Positive Predictive Value) of 50% for the discordant class (Class 1), and an overall Accuracy of 76%. SHAP analysis revealed that the most important predictors of discordance were female gender, advanced age, and the presence of diabetes, indicating that the prediction was influenced by the combination of these factors rather than a single one.

Conclusion: This exploratory study demonstrates the potential value of explainable AI for deciphering complex clinical problems such as MPS/ECG discordance. Our multivariate models identified a potential patient profile associated with of MPS-/ECG+ discordance, characterized by the synergy of clinical factors. These preliminary results suggest that ECG abnormalities in the absence of a perfusion deficit might reflect an underlying pathology (e.g. microvascular disease) rather than a simple false positive, warranting further prospective validation.

Keywords: coronary artery disease, myocardial perfusion SPECT mps, explainable AI, SHAP in diagnostic discordance, risk stratification, microvascular disease

Introduction

Cardiac function evaluation and coronary artery disease (CAD) diagnosis classically rely on the electrocardiogram (ECG). The combination of a stress ECG and myocardial perfusion SPECT (MPS) represents a robust technique for CAD diagnosis.¹ Although MPS has generally been shown to be superior to exercise ECG in terms of sensitivity and specificity for the diagnosis of obstructive CAD,² discrepancies between the two modalities are frequently encountered in clinical practice.³

A particularly challenging and intriguing clinical scenario is that of patients presenting with a positive stress ECG (suggesting ischemia) but a negative MPS (indicating no significant perfusion defect). This profile, often described as “paradoxical” or an ECG “false positive”, poses a real diagnostic challenge. It may lead to unnecessary invasive investigations or, more worryingly, conceal distinct pathophysiological mechanisms.³ In a previous study conducted at our own centre, we observed that approximately 13% of patients presented with this discordance (positive ECG and negative MPS), highlighting the prevalence of the clinical problem.⁴ Furthermore, balanced ischemia, where multiple vessels are affected equally, can lead to false-negative MPS results despite positive ECG findings.⁵ Indeed, recent research suggests that, in these patients, the electrical abnormality on the ECG is not a random error but may reflect coronary microvascular dysfunction (CMD),⁶ a pathology that traditional MPS does not always detect.^{3,7} The management of these potentially at-risk patients, despite a normal MPS, requires precise identification of their profile.

Initial bivariate analyses often struggle to identify clear predictors for such complex phenomena, especially in small cohorts.⁸ Facing this complexity, Artificial Intelligence (AI) and Machine Learning (ML) offer exploratory perspectives to model non-linear relationships and multivariate interactions.⁸ These techniques are capable of modelling non-linear relationships and multivariate interactions among clinical, demographic, and imaging factors, thereby surpassing the limitations of conventional statistics to enhance prognostic and diagnostic value.^{9,10} This is consistent with the rapidly evolving literature demonstrating the broad utility of machine learning applications in advanced ECG interpretation.¹¹ Recent systematic reviews confirm the role of AI in improving the sensitivity and specificity of stress tests and reducing false-positive rates.¹² This momentum is further evidenced by very recent advancements where machine learning has proven highly effective not only in predicting prognosis using SPECT data,¹³ but also across other advanced cardiac imaging modalities such as AI-powered contrast-free cardiovascular magnetic resonance for myocardial infarction.¹⁴ Furthermore, interpretable AI models (Explainable AI - XAI) have become crucial for clinical validation, allowing researchers to identify and understand the underlying factors driving model decisions.⁸

Our underlying hypothesis is that discordant cases (MPS-/ECG+) do not merely represent random false-positive ECGs, but rather reflect a distinct, multifactorial clinical profile—potentially linked to microvascular dysfunction—that AI algorithms can identify through complex, non-linear feature interactions.

The objective of this exploratory study is twofold: (1) to compare the clinical characteristics of MPS-negative patients based on their stress ECG results using bivariate analysis; and (2) to apply explainable AI models (Random Forest, XGBoost and SHAP) to explore potential synergistic and non-linear interactions among clinical factors that may characterize the MPS-/ECG+ discordance profile.

Materials and Methods

Study Population

This monocentric retrospective study was conducted on a database of patients referred for stress myocardial perfusion SPECT (MPS) between January 2017 and December 2022, at the Department of Nuclear Medicine, CHU MOHAMMED VI in Marrakech. Inclusion criteria were:

- age over 18,
- performance of an MPS coupled with an exercise test on a treadmill or cycle ergometer, and
- an MPS result interpreted as negative for reversible or fixed ischaemia and necrosis. Patients who had undergone pharmacological testing were excluded.

Of an initial cohort of 180 patients, 86 met these criteria and formed the final study population. This cohort was divided into two groups on the basis of the electrical stress test result:

- Discordant group (Class 1): Patients with negative MPS and positive stress ECG, defined by the occurrence of horizontal or descending ST segment sub-shift ≥ 1 mm. (n=19)

- Concordant Group (Class 0): Patients with negative MPS and a stress ECG ultimately classified clinically as negative (n=67). It is important to note that this real-world group includes some initially “inconclusive” ECG tracings that were retrospectively deemed non-ischemic and clinically negative by the interpreting physicians in light of the normal MPS results.

This stratification allowed for the isolation of the clinical subgroup of interest (MPS-/ECG+), which represented 22.1% of the final cohort.

Data Collection and Preparation

Initially, a set of 40 variables was extracted from the computerized medical records. However, 20 variables representing scintigraphic perfusion parameters were strictly normal for all patients (due to the inclusion criteria of negative MPS) and had zero variance. Consequently, these 20 non-informative variables were excluded prior to modelling. The machine learning models were trained on the remaining clinical, demographic, and echocardiographic variables. These variables included:

- Demographic data: Age, Sex.
- Cardiovascular risk factors: Presence/absence of diabetes, High blood pressure (HBP), active or cessation of smoking, dyslipidaemia, obesity or overweight, sedentary lifestyle, and menopausal status.
- Clinical history: Previous myocardial infarction (MI) or acute coronary syndrome (ACS), revascularisation by coronary angioplasty or coronary artery bypass grafting, known cardiomyopathy (CMI), and conduction disorders (bundle branch block, atrioventricular block).
- Cardiac imaging data: Left ventricular ejection fraction (LVEF) from the most recent transthoracic echocardiogram.

Categorical data (eg. gender) was digitally encoded (One-Hot Encoding) to give two new variables (Sex_F and Sex_M) for modelling in machine learning.

Conventional Statistical Analysis

A bivariate statistical analysis was performed to compare the characteristics of the two groups: Discordant Group (MPS-/ECG+) and Concordant Group (MPS-/ECG-).

Continuous variables were presented as mean $\pm$ standard deviation and compared using Student's *t*-test for independent samples.

Categorical variables were presented as numbers (percentage) and compared using Pearson's Chi-squared test or Fisher's exact test, depending on the numbers involved.

The statistical significance threshold was set at $p < 0.05$. Analyses were performed using the *scipy.stats* library in Python.

Machine Learning Approach

Data Preprocessing, Set Splitting and Imbalance Management

All 86 patients were divided into a training set (70%, n=60) and a test set (30%, n=26). The split was stratified to preserve the proportion of discordant cases (22.1%) in both subsets, with the random seed fixed (*random_state*=42) to ensure reproducibility. Prior to modelling, data preprocessing included the handling of rare missing values for continuous variables via mean imputation, and the transformation of categorical variables (eg., gender) using one-hot encoding. No complex algorithmic feature selection was performed beyond the initial exclusion of zero-variance scintigraphic variables.

Given the inherent class imbalance (13 discordant vs. 47 concordant cases in the training set), the Synthetic Minority Over-sampling Technique (SMOTE) was applied. Crucially, SMOTE was strictly limited to the training set to prevent data leakage. It generated exactly 34 synthetic samples of the minority class, resulting in a perfectly balanced training set

of 94 samples (47 per class). The test set (n=26) remained entirely real and unmodified to ensure that performance metrics were not artificially inflated.

Furthermore, to mitigate the severe risk of overfitting on this small dataset, we deliberately avoided exhaustive automated hyperparameter tuning (such as GridSearch). Instead, models were trained using manually constrained, conservative hyperparameters (eg., for XGBoost: max_depth = 3, subsample = 0.8; for Random Forest: max_depth = 10, min_samples_leaf = 2). The classification threshold was maintained at the standard default value of 0.5. A comprehensive breakdown of all variables and their specific preprocessing steps is detailed in Table 1.

Classification Algorithms and Training

Two supervised learning algorithms of the ensemble learning classification type were trained on the SMOTE-processed dataset and compared:

- Random Forest (RF): A model based on the aggregation of multiple decision trees, recognised for its robustness and its low sensitivity to overfitting.

Table 1 Features of the MPS-Negative Population According to Stress ECG Result

Features	Group MPS-/ECG- (n=67)	Group MPS-/ECG+ (n=19)	P-value
Age (years), mean $\pm$ SD	60.1 $\pm$ 10.4	58.6 $\pm$ 6.0	0.560
Gender Female, n (%)	29 (43.3%)	12 (63.2%)	0.269
Gender Male, n (%)	38 (56.7%)	7 (36.8%)	0.269
Diabetes, n (%)	25 (37.3%)	6 (31.6%)	0.747
HBP, n (%)	28 (41.8%)	9 (47.4%)	0.987
Smoking, n (%)	8 (11.9%)	2 (10.5%)	1.000
LVEF (%), mean $\pm$ SD	59.8 $\pm$ 5.7	59.7 $\pm$ 3.3	0.947
Age Mentioned as RF	10 (14.9%)	5 (26.3%)	0.469
Dyslipidaemia	10 (14.9%)	3 (15.8%)	1.000
Obesity	5 (7.5%)	0 (0.0%)	0.502
Sedentary	9 (13.4%)	1 (5.3%)	0.565
Menopause	6 (9.0%)	2 (10.5%)	1.000
Renal failure	5 (7.5%)	0 (0.0%)	0.502
BBB, n (%)	7 (10.4%)	1 (5.3%)	0.769
MI/ACS	6 (9.0%)	5 (26.3%)	0.107
Coronary stent/angioplasty	6 (9.0%)	5 (26.3%)	0.107
CABG	2 (3.0%)	0 (0.0%)	1.000
ICM	5 (7.5%)	1 (5.3%)	1.000
LVG	5 (7.5%)	1 (5.3%)	1.000
AV block	1 (1.5%)	0 (0.0%)	1.000

Abbreviations: SD, Standard deviation; HBP, High blood pressure; LVEF, Left ventricular ejection fraction; RF, Risk factor; BBB, bundle branch block; MI, Myocardial infarction; ACS, Acute Coronary Syndrome; CABG, Coronary Artery Bypass Grafting; AV block, Atrioventricular block; LVG, Left ventricular hypertrophy; ICM, ischaemic cardiomyopathy.

- XGBoost (Extreme Gradient Boosting): A model based on the principle of boosting, where trees are built sequentially to correct the errors of previous ones, often at the cutting edge of performance on tabular data.

Performance Evaluation

The performance of models was evaluated on the unmodified test set using several classic classification metrics:

- Accuracy: Overall proportion of correct predictions
- Precision (Positive Predictive Value): Proportion of true positive predictions out of all positive predictions for the discordant class (Class 1).
- Recall (Sensitivity): Proportion of true positive predictions out of all actual discordant cases.
- F1-Score: Harmonic mean of Precision and Recall.
- Area under the ROC curve (AUC), which measures the model's overall discrimination ability.

Model Interpretability (Explainable AI - XAI)

To understand the decision-making mechanisms of the models, we used the SHAP (SHapley Additive exPlanations) algorithm. This method, derived from game theory, assigns each patient variable an impact value (SHAP value) on the final prediction. SHAP graphs were generated to visualise:

- The overall importance of the variables for each model.
- The impact (direction and magnitude) of the value of each variable on the probability of predicting the discordant class.
- Detailed explanation of the prediction for individual cases.

All machine learning analyses were carried out in Python using the scikit-learn, imblearn, xgboost and SHAP libraries.

Results

Descriptive and Statistical Analysis of the Cohort

The study cohort included 86 patients with negative myocardial perfusion SPECT (MPS). This population was subdivided into two groups based on the stress ECG result, as illustrated in [Figure 1](#):

- Discordant Group (MPS-/ECG+): 19 patients (ie., 22.1%).
- Concordant Group (MPS-/ECG-): 67 patients (ie., 77.9%).

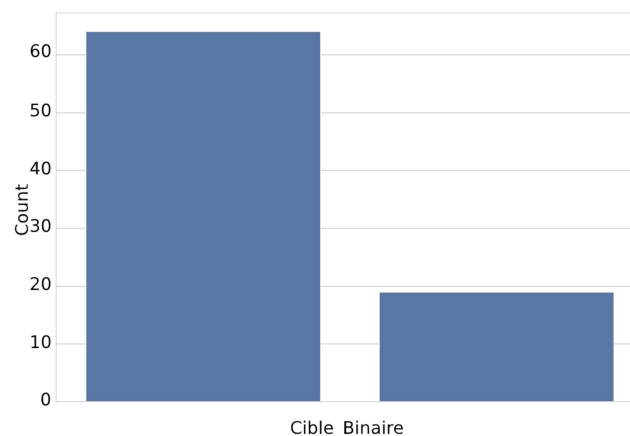


Figure 1 Distribution of Classes.

Table 2 Comparison of Model Performance on the Test Set (n=26)

Model	Accuracy	Precision (Class 1)	Recall (Class 1)	F1-Score (Class 1)	AUC
Random Forest	0.68	0.38	0.50	0.43	0.61
XGBoost	0.76	0.50	0.67	0.57	0.78

Bivariate statistical analysis was conducted to identify any baseline differences between these two groups (Table 1).

As shown in Table 1, the bivariate statistical analysis did not reveal any variable significantly associated with discordance (all p-values > 0.05). The main clinical and demographic characteristics (age, sex, diabetes) were similar between the two groups, suggesting that the MPS-/ECG+ discordance phenomenon cannot be explained by a single, isolated factor.

Comparative Performance of Artificial Intelligence Models

In the absence of a univariate signal, a multivariate approach was adopted. Two models, Random Forest (RF) and XGBoost, were trained and evaluated. The XGBoost model demonstrated superior performance on all metrics, as summarised in Table 2.

The most notable improvement concerns the key metrics for the minority (discordant) class:

- The Recall increased from 50% (RF) to 67% (XGBoost), demonstrating a better identification of true discordant cases.
- The AUC rose from 0.61 (a weak model) to 0.78 (a moderately good and informative model), illustrating a globally superior discrimination capability.

The confusion matrices (Figure 2A and B) and comparative ROC curves (Figure 3) visually illustrate the superiority of the XGBoost model, which manages to identify discordant cases better while maintaining a low false positive rate:

- Random Forest: identified 3 True Positives and 3 False Negatives (Recall de 3/6=50%).
- XGBoost: identified 4 True Positives and 2 False Negatives (Recall de 4/6=67%).

XGBoost correctly identified one additional discordant patient compared to Random Forest (4 true positives versus 3, out of 6 actual discordant cases in the test set). The AUC increased from 0.61 (Random Forest) to 0.78 (XGBoost), demonstrating improved discriminative ability.

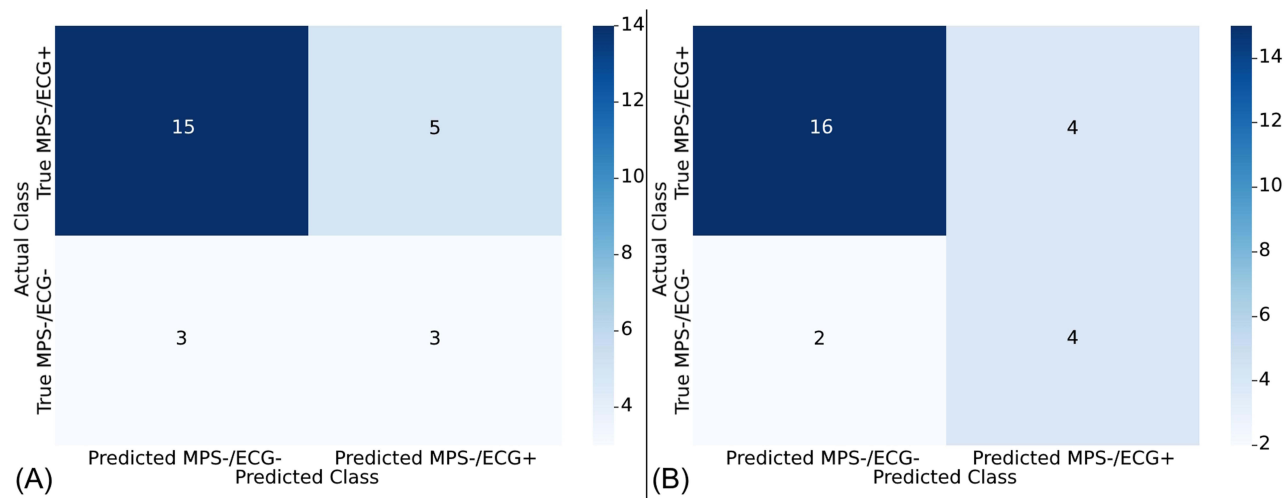


Figure 2 Confusion Matrices (A) Random Forest, (B) XGBoost.

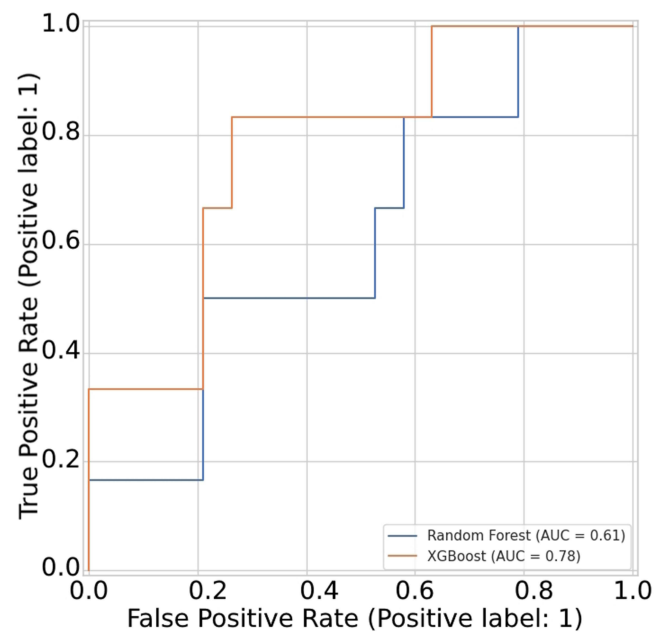


Figure 3 Comparative ROC curves for the two models.

Interpretability of Models and Identification of Key Factors

Analysis of the importance of the variables revealed a remarkable convergence between the two models. As shown in [Figure 4A](#) (Top 20 Random Forest) and [Figure 4B](#) (Top 20 XGBoost), both algorithms rank diabetes, gender, age and LVEF as the most influential factors in their decision making, albeit in a slightly different order.

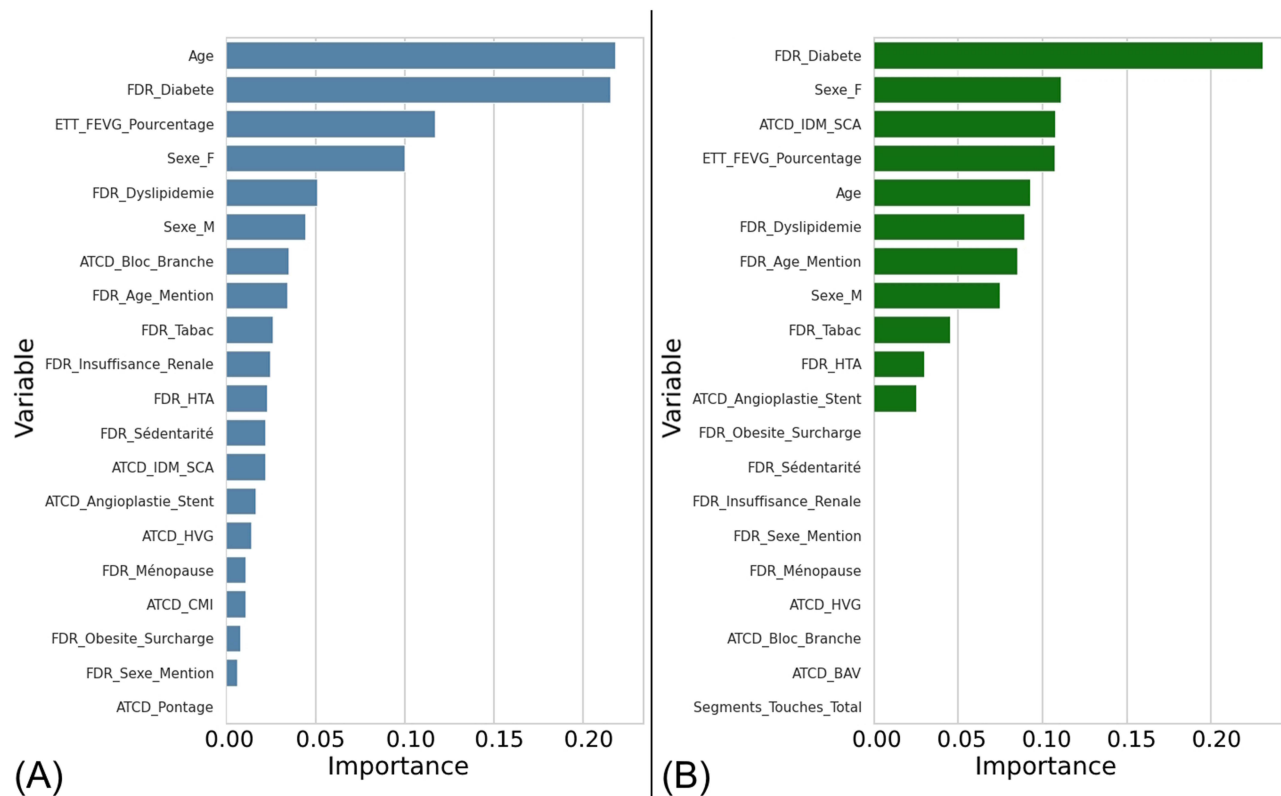


Figure 4 Importance of variables for RF (A) and XGBoost (B).

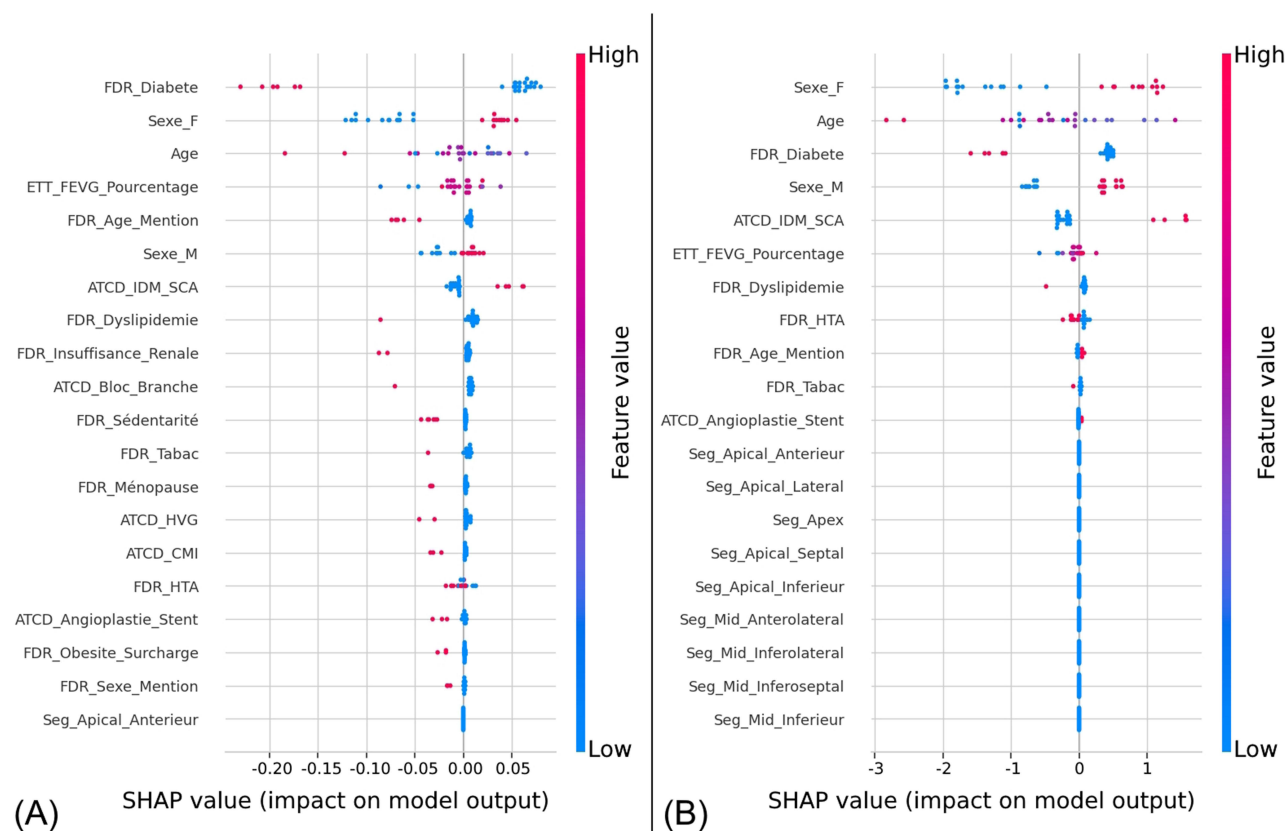


Figure 5 Detailed SHAP graphs for Random Forest (A) and XGBoost (B).

An important point of divergence between the two models concerns the History of Myocardial Infarction or Acute Coronary Syndrome (ATCD_IDM_SCA). This variable ranks third among the most important factors in the XGBoost model, while being much less influential in Random Forest (Figure 4A and B). The SHAP analysis for XGBoost (Figure 5B), provides insight into this difference.

The impact of ATCD_IDM_SCA: Although ranked as important, examining the SHAP dispersion shows that the presence of a history (red dots) is largely distributed on both sides of the zero line, but with a concentration of points that pulls towards non-discordance (negative SHAP). This suggests that the non-presence of a history (blue dots, low feature value) is actually a factor pushing the prediction towards discordance for some patients.

This complexity illustrates one of the benefits of AI: the higher-performing XGBoost model (AUC 0.78) was able to identify an indirect or contextual relationship with ATCD_IDM_SCA. We can hypothesize that the presence of a visible infarction may have been the cause of the ECG abnormality or may have led to easily identifiable structural changes, making the explanation for the MPS-/ECG+ discordance less “paradoxical.” Consequently, it is the patients without a major history who represent the greatest diagnostic challenge and constitute the true subgroup of interest for the model.

The full convergence and directional impact are confirmed by the analysis of the SHAP values, presented in the detailed summary plots (Figure 5A) SHAP Random Forest: MPS-/ECG+ and Figure 5B) SHAP XGBoost: MPS-/ECG+).

By focusing on the forces that push the prediction toward the positive class (SHAP value > 0, to the right of the central line), the models clearly identify the distinctive profile of the discordant patient:

- The presence of diabetes (red dots in the FDR_Diabete row) and female gender (red dots in the Sexe_F row) are the strongest positive predictors of discordance.
- Older age (Age) also significantly tends to increase the probability of discordance.

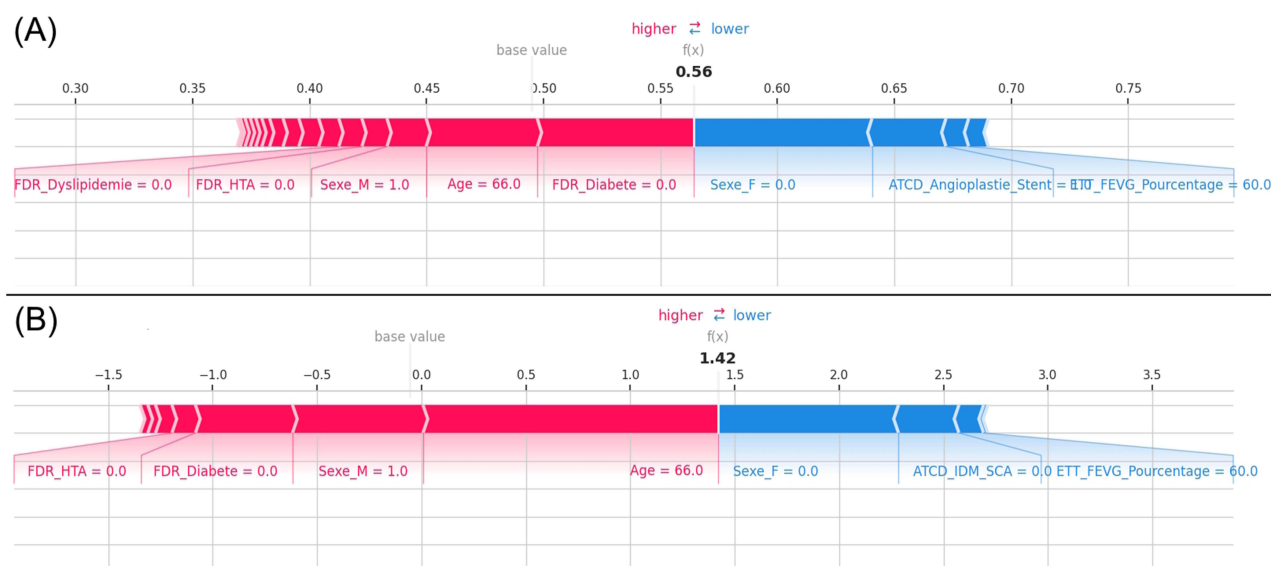


Figure 6 Illustrative example of the model explanation process for a single discordant case (patient n°25), illustrated by a SHAP Force Plots graph. **(A)** Random Forest, **(B)** XGBoost.

- Conversely, a high LVEF (ETT_FEVG_Pourcentage) (blue dots) is consistently associated with a lower probability of discordance (negative SHAP values).

This convergence between the two distinct AI models suggests that MPS-/ECG+ discordance is unlikely to be driven by a single factor, but rather highlights potentially important combinations of variables. Given the limited sample size, these SHAP outputs do not confirm a stable or definitive causal pattern, but rather generate strong exploratory hypotheses regarding the synergy of specific clinical factors (female gender, diabetes, and advanced age).

The analysis of an individual discordant case (test patient n°25) further illustrates the reasoning of the models, as shown in Figure 6, where Random Forest and XGBoost predictions are compared. For this 66-year-old non-diabetic man with an LVEF of 60%, both models correctly predicted his discordant nature.

- Random Forest Prediction (Figure 6A): The prediction relied on a slightly different combination, with Age and the absence of dyslipidemia contributing to the positive prediction, while Sexe_F and LVEF remained strong negative (protective) influences.
- XGBoost Prediction (Figure 6B): The advanced age (Age = 66.0) was the primary factor driving the prediction toward discordance (red force), while the preserved LVEF (60.0%) and the absence of female sex (Sexe_F = 0.0) acted as moderating, protective factors (blue forces).

This detailed patient-level analysis shows not only that models can capture complex profiles, but also that the choice of model can influence the interpretation of the relative weights of each factor for a given decision.

Discussion

The most striking result of our study is the paradox between the inability of bivariate analyses to identify significant predictors and the ability of AI models to construct an informative classifier from the same data. This finding does not represent a failure of analysis, but on the contrary reinforces the interest of the AI approach for complex clinical problems. It hypothesizes that MPS-/ECG+ discordance is a multifactorial phenomenon, whose signature lies not in a single dominant factor, but in a subtle interaction of several variables.

Characterization of the Discordant Profile by AI

A key finding of this study is the apparent paradox between the lack of statistical significance in the initial bivariate analysis (Table 1) and the high importance assigned to female sex, advanced age, and diabetes by the AI models (SHAP analysis). While these factors failed to discriminate the groups when analysed in isolation ($p > 0.05$), the ensemble AI models (XGBoost and Random Forest) successfully captured complex, non-linear synergistic interactions among them. Consequently, rather than a single dominant factor, it is the combination of these specific clinical variables that exploratory AI suggests as a potential profile associated with MPS-/ECG+ discordance.

Female gender is a well-known factor for stress test false positives,^{15,16} but our model identifies it as the most important factor in a multivariate analysis. Recently, AI applied to the analysis of stress ECG (ExECG) has also highlighted sex as a primary predictive variable,¹⁷ reinforcing the notion that the ECG signal in women presents distinctive characteristics that AI is capable of identifying.

The role of diabetes is also fundamental, as it is an established risk factor for coronary microvascular disease (CMD),⁷ a condition that can induce electrical abnormalities on the stress ECG (ischemia) without a visible perfusion deficit on MPS (which assesses perfusion at the macroscopic level).¹⁸ The study by Sinha et al (2024) confirms that in patients with a positive ExECG but normal coronary angiography, microvascular function is altered, justifying increased vigilance for this phenotype.⁶ Our model, by identifying this synergy of factors (Female, Diabetes, Age), shows promising potential as an exploratory tool for this CMD.³ Recent systematic reviews further confirm that Machine Learning can be utilized to better manage false positives in stress testing.¹²

Furthermore, the observation that a high Left Ventricular Ejection Fraction (LVEF) is associated with a lower probability of discordance is a relevant finding, as it suggests that the MPS-/ECG+ discordance is a phenomenon that manifests before the onset of significant functional alteration, or that the ECG signal is less perturbed in the absence of major ventricular remodeling.¹⁹

This profile identification is further refined by the complex influence of the History of Myocardial Infarction or Acute Coronary Syndrome (ATCD_IDM_SCA). While the presence of these antecedents might be expected to explain the ECG abnormality, the XGBoost model assigns high importance to this factor, but the SHAP analysis suggests that the absence of a major MI/ACS history is actually linked to the discordant profile for some patients. This highlights that the MPS-/ECG+ discordance is truly paradoxical in patients without macrovascular disease history. This is consistent with the hypothesis that this phenotype primarily reflects a functional or microvascular disorder,⁶ rather than a classic post-infarction structural defect. The strong importance of ATCD_IDM_SCA in XGBoost, coupled with the influence of Age and LVEF, strongly points toward the role of cardiac structural changes (eg., hypertrophy and fibrosis) as modulators of the ECG signal in this complex patient group.¹⁹

Superiority of Boosting Learning and Interpretability

The significant improvement in performance observed with the XGBoost model (AUC of 0.78) compared with the Random Forest (AUC of 0.61) indicates that the relationships between the variables are particularly complex and non-linear. A powerful algorithm based on Gradient Boosting is better suited to modelling second-order interactions and the synergy of these subtle clinical factors.

The use of interpretability tools like SHAP is purely advantageous. In a domain where clinical decision-making must be transparent, SHAP allowed us to move from a simple classifier to a tool for generating clinical hypotheses.

The analysis of an individual case illustrates how the two models arrive at a prediction. For test patient 25 (a 66-year-old non-diabetic man with an LVEF of 60%), both models correctly predicted that he belonged to the discordant group. However, their “reasoning” differed. While Random Forest attributed the strongest impact to the patient’s sex, XGBoost, the best performing model, identified advanced age as the main factor driving the prediction towards discordance. This detailed analysis at patient level shows not only that models can capture complex profiles, but also that the choice of model can influence the interpretation of the relative weights of each factor for a given decision.

Clinical Implications and Pathophysiological Mechanisms

These results generate a strong clinical hypothesis: The positivity of the stress ECG in these patients is not a random error, but the reflection of a particular pathophysiological substrate.⁶ Several potentially interrelated mechanisms could be involved, primarily:

- **Microvascular disease:** A common complication in diabetic patients, this can cause ischaemia that cannot be detected by conventional macroscopic perfusion imaging (MPS), which assesses perfusion at macroscopic level. However, as we did not directly assess microvascular function, this remains a speculative, albeit highly plausible, hypothesis supported by recent literature.^{7,18}
- **Cardiac structural changes:** Age and hypertension (often associated with diabetes) are major risk factors for cardiac structural and functional changes, including left ventricular hypertrophy (LVH) and fibrosis, altering repolarisation during exercise.^{19,20}
- **Hormonal and anatomical factors:** The greater prevalence of this discrepancy in women is a well-known fact, often attributed to hormonal influences such as high oestrogen levels, or anatomical differences such as breast attenuation.^{16,21,22} Our model quantified and ranked this factor as the most important.

It is imperative to note that SHAP values explain the predictive behaviour of the machine learning model itself, rather than proving direct causal or pathophysiological mechanisms. Therefore, the patterns identified by our models primarily serve to generate clinical hypotheses.

Limitations and Perspectives

It is important to recognize several major limitations of our work. Firstly, the overall cohort size is small, particularly the number of patients in the discordant group (n=19). This severely limits the statistical power of the initial analyses, which most likely explains the lack of significance in the conventional statistical tests, and heightens the risk of overfitting in the AI models.

In the used method, our evaluation was based on a single 70/30 split between test and train data, instead of resorting to more rigorous internal techniques such as k-fold cross-validation or bootstrapping. Moreover, no analysis for the model calibration was performed. For the reported performance metrics such as AUC and Sensitivity, these indicators are solely point estimates. And this due to the small sample size (n=26). Furthermore, we did not calculate confidence intervals, which resulted in significant statistical uncertainty, meaning that the reliability of the models is limited.

This study also has limitations related to its single-centre, retrospective nature, as well as the lack of external validation. In addition, we did not integrate an appropriate conventional multivariate comparison model (such as logistic regression) to establish a baseline. In the absence of such a reference, the current results serve to demonstrate AI's ability to model complex, non-linear interactions, rather than to assert any superiority over established conventional methods.

From a clinical perspective, despite our hypothesis that microvascular dysfunction (CMD) might explain the MPS-/ECG+ discordance observed in our cohort, the absence of direct and objective assessment of microvascular function (for example, via measurement of coronary flow reserve) represents a major limitation. Consequently, the link established between this clinical profile and CMD remains merely a hypothesis and requires prospective validation.

For these reasons, the results should be interpreted with caution. Furthermore, as the models presented in this study should be regarded strictly as exploratory tools for generating hypotheses, they are not currently reliable or effective enough to be used directly for diagnostic purposes in routine clinical practice.

Conclusion

This exploratory study demonstrates that artificial intelligence, coupled with interpretability tools such as SHAP, is proving to be a promising approach and a powerful method for generating clinical hypotheses in complex scenarios where bivariate analyses are limited. We have suggested a potential multifactorial profile of patients at risk of MPS-/ECG + discordance, which appears to be centred on female gender, age, and diabetes.

The consistency of these preliminary findings with the literature strongly justifies further research in this context, through the conduct of multicentre studies on significantly larger cohorts. Furthermore, confirming the stability of these results requires rigorous external validation to establish their true clinical value. The development of such reliable and predictive tools could assist risk stratification by clinicians and guide the management of patients presenting this particular clinical profile.

Ethical Statement

The data used in this retrospective study were derived exclusively from the medical records of patients in the Department of Nuclear Medicine at the Mohammed VI University Hospital in Marrakesh. The study protocol, which was strictly observational and non-interventional, had no impact on the diagnosis, therapeutic strategies or clinical management of patients. Patient data was irreversibly anonymised in its entirety before any analysis, in order to ensure strict confidentiality and the utmost protection of patients' privacy. Consequently, the dataset used in this study contains no "personal data" as defined by Moroccan Law No. 09-08 relating to the protection of individuals concerning the processing of personal data. In this case, formal approval from an ethics committee was not deemed necessary under institutional guidelines for this specific type of research.

Disclosure

All authors declare that they have no conflicts of interest to this work.

References

1. Kraen M, Akil S, Heden B, et al. The incremental value of exercise ECG to myocardial perfusion SPECT for prediction of cardiac events. *Eur Heart J*. 2022;43(Supplement_2):ehac544.1154. doi:10.1093/eurheartj/ehac544.1154
2. Bokhari S, Shahzad A, Bergmann SR. Superiority of exercise myocardial perfusion imaging compared with the exercise ECG in the diagnosis of coronary artery disease. *Coronary Artery Disease*. 2008;19(6):399–404. doi:10.1097/MCA.0b013e3283021ab4
3. Qamruddin S. False-positive stress echocardiograms: a continuing challenge. *Ochsner J*. 2016;16(3):277–279. doi:10.1016/j.ahj.2011.11.022
4. Elmaadaoui Y, Belaguid A, Bsis MA, Matrane A. Study of the sensitivity and specificity of mps and ecg in detecting myocardial ischemia in the inferior wall. In: Ezziyyani M, Kacprzyk J, Balas VE, eds. *International Conference on Advanced Intelligent Systems for Sustainable Development (AI2SD 2024)*. Vol 1402. Lecture Notes in Networks and Systems. Springer Nature Switzerland; 2025:732–740. doi:10.1007/978-3-031-91334-1_64.
5. Baqi A, Ahmed I, Nagher B. Multi vessel coronary artery disease presenting as a false negative myocardial perfusion imaging and true positive exercise tolerance test: a case of balanced ischemia. *Cureus*. doi:10.7759/cureus.11321
6. Sinha A, Dutta U, Demir OM, et al. Rethinking false positive exercise electrocardiographic stress tests by assessing coronary microvascular function. *J Am Coll Cardiol*. 2024;83(2):291–299. doi:10.1016/j.jacc.2023.10.034
7. Swaraj S, Kott K, Vernon S, Figtree G. 738 microvascular dysfunction: assessment of risk factors in patients with positive stress tests but no obstructive epicardial coronary artery disease. *Heart Lung & Circulation*. 2020;29:S368. doi:10.1016/j.hlc.2020.09.745
8. Shameer K, Johnson KW, Glicksberg BS, Dudley JT, Sengupta PP. Machine learning in cardiovascular medicine: are we there yet? *Heart*. 2018;104(14):1156–1164. doi:10.1136/heartjnl-2017-311198
9. Shrestha S, Sengupta PP. Machine learning for nuclear cardiology: the way forward. *J Nucl Cardiol*. 2019;26(5):1755–1758. doi:10.1007/s12350-018-1284-x
10. Seetharam K, Min JK. Artificial intelligence and machine learning in cardiovascular imaging. *Methodist Debaquey Cardiovasc*. 2020;16(4):263. doi:10.14797/mdcj-16-4-263
11. He Y, Zhou Y, Qian Y, et al. Cardioattentionnet: advancing ECG beat characterization with a high-accuracy and portable deep learning model. *Front Cardiovasc Med*. 2025;11:1473482. doi:10.3389/fcvm.2024.1473482
12. Hadida Barzilai D, Cohen-Shelly M, Sorin V, et al. Machine learning in cardiac stress test interpretation: a systematic review. *Eur Heart J Digit Health*. 2024;5(4):401–408. doi:10.1093/ehjdh/ztae027
13. Cicek V, Cikirikci EH, Babaoğlu M, et al. Machine learning for prognostic prediction in coronary artery disease with SPECT data: a systematic review and meta-analysis. *EJNMMI Res*. 2024;14(1):117. doi:10.1186/s13550-024-01179-2
14. Cicek V, Bagci U. AI-powered contrast-free cardiovascular magnetic resonance imaging for myocardial infarction. *Front Cardiovasc Med*. 2024;11:1457498. doi:10.3389/fcvm.2024.1457498
15. Fitzgerald BT, Scalia WM, Scalia GM. Female false positive exercise stress ecg testing – fact versus fiction. *Heart Lung & Circulation*. 2019;28(5):735–741. doi:10.1016/j.hlc.2018.02.010
16. Judelson DR. Examining the gender bias in evaluating coronary disease in women. *Medscape Womens Health*. 1997;2(2):5.
17. Banerjee A, Salian RS, Vemulapalli HS, et al. Enhancement of stress ecg performance with machine learning. *JACC: Advances*. 2025;4(10):102141. doi:10.1016/j.jacadv.2025.102141
18. Levy BI, Schiffrin EL, Mourad JJ, et al. Impaired tissue perfusion: a pathology common to hypertension, obesity, and diabetes mellitus. *Circulation*. 2008;118(9):968–976. doi:10.1161/CIRCULATIONAHA.107.763730
19. Saheera S, Krishnamurthy P. Cardiovascular changes associated with hypertensive heart disease and aging. *Cell Transplant*. 2020;29:096368972092083. doi:10.1177/0963689720920830
20. Zhang X, Wang X, Li L, Zhang G, Gao Y, Cui J. An analysis of factors influencing electrocardiogram stress test for detecting coronary heart disease. *Chin Med J*. 1999;112(7):590–592.

21. Clark PI, Glasser SP, Lyman GH, Krug-Fite J, Root A. Relation of results of exercise stress tests in young women to phases of the menstrual cycle. *Am J Cardiol.* 1988;61(1):197–199. doi:10.1016/0002-9149(88)91334-3
22. Mehta PK, Isiadinso I, Wann LS. *Stress Testing in Women*. Thomas GS, Wann LS, Ellestad MH. eds..6th. Ellestad's Stress Testing Oxford University Press New York. 2018. 473–491. doi:10.1093/med/9780190225483.003.0024

Medical Devices: Evidence and Research

Dovepress
Taylor & Francis Group

Publish your work in this journal

Medical Devices: Evidence and Research is an international, peer-reviewed, open access journal that focuses on the evidence, technology, research, and expert opinion supporting the use and application of medical devices in the diagnosis, monitoring, treatment and management of clinical conditions and physiological processes. The identification of novel devices and optimal use of existing devices which will lead to improved clinical outcomes and more effective patient management and safety is a key feature of the journal. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/medical-devices-evidence-and-research-journal>