ESC European Society of Cardiology
European Heart Journal - Digital Health (2025) **6**, 1113–1123
https://doi.org/10.1093/ehjdh/ztaf092

ORIGINAL ARTICLE

Prognostic stratification of familial hypercholesterolaemia patients using AI algorithms: a gender-specific approach

Alberto Zamora [1,2,3,4,*], Luis Masana [4,5,6,*], Fernando Civeira [7,8], Daiana Ibarretxe [4,5], Marta Fanlo-Maresma [9,10], Alex Vila [11], Manuel Suárez Tembra [12], Victoria Marco-Benedí [13,14], Luis A. Alvarez-Sala-Walther [15,16], and Miguel Camacho-Ruiz [17,*]; on behalf of the National Registry of the Spanish Society of Arteriosclerosis

[1]Instituto de Investigación Biomédica de Girona Dr. Josep Trueta (IDIBGI), Salt, Girona 17190, Spain; [2]Digital Health Group, Corporació de Salut del Maresme I la Selva, Calella, Barcelona 08370, Spain; [3]National Center for Cardiovascular Research (CIBERCV), Spain; [4]Catalonia Lipid Units Network (XULA), Catalonia, Spain; [5]Unitat de Medicina Vascular i Metabolisme, Hospital Universitari Sant Joan, Reus, Tarragona, España; [6]Unitat de Reçerca en Lípids I Arteriosclerosi, Universitat Rovira I Virgili, IISPV, Reus, Tarragona, España; [7]Centro de Investigación Biomédica en Red de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), Calle Monforte de Lemos, 3-5, Pabellón 11, Planta baja, 28029 Madrid, España; [8]Hospital Universitario Miguel Servet, IIS Aragón, CIBER Cardiovascular, Universidad de Zaragoza, Zaragoza, Spain; [9]CIBER Fisiopatología de la Obesidad y Nutrición (CIBEROBN), Instituto de Salud Carlos III (ISCIII), Madrid, Spain; [10]Lipid and Vascular RiskUnit, Internal Medicine Service, Hospital Universitario de Bellvitge, Instituto de Investigación Biomédica de Bellvitge (Idibell), Hospitalet de Llobregat, Spain; [11]Unidad de Lípidos, Hospital de Figueres, Fundació Salut Empordà, Figueres, Spain; [12]Unidad de Lípidos Hospital San Rafael, A Coruña, Calle Monforte de Lemos, 3-5, Pabellón 11, Planta baja, 28029 Madrid, Spain; [13]Hospital Universitario Miguel Servet, IIS Aragón, CIBERCV, Zaragoza, Calle Monforte de Lemos, 3-5, Pabellón 11, Planta baja, 28029 Madrid, Spain; [14]Universidad de Zaragoza, Zaragoza, Spain; [15]Internal Medicine, Hospital General Universitario Gregorio Marañon, IiSGM Calle Monforte de Lemos, 3-5, Pabellón 11, Planta baja, 28029 Madrid, Spain; [16]Universidad Complutense Madrid, Madrid 28007, Spain; and [17]Universidad Complutense de Madrid, Faculty of Statistical Studies, Madrid, Spain

Received 12 February 2025; revised 3 May 2025; online publish-ahead-of-print 26 August 2025

Aims

Familial hypercholesterolaemia (FH) is the most prevalent autosomal dominant disorder, affecting about 1 in 200–250 individuals. It is the leading cause of early and aggressive coronary artery disease.

Methods and results

We analysed patients with genetically confirmed FH or a score >8 on the Dutch Lipid Clinics Network criteria from the National Registry of the Spanish Atherosclerosis Society, including individuals enrolled from January 2010 to December 2017. The model utilized a dataset incorporating family history, clinical characteristics, laboratory results, genetic data, imaging studies, and lipid-lowering treatment details. Eighty per cent of the population was allocated for training the AI algorithm and 20% was used for testing. A Histogram-based Gradient Boosting Classification Tree was used. The stability of the AI system was assessed using *K*-fold cross-validation. Shapley additive explanations methodology analysed the influence of different variables by sex. Youden's *J* statistic established the optimal cut-off point. A total of 1764 patients were included (51.8% women), among whom 264 experienced major adverse cardiovascular events (MACEs), with 8% being women. The final model incorporated 82 variables, achieving metrics of precision for MACE accuracy (0.92), recall (0.89), F1-score (0.91), and receiver operating characteristic (0.88; 95% confidence interval, 0.85–0.90). In the model, age, gamma-glutamyl transferase levels, and subclinical disease significantly impacted risk for women, while year of birth, age at initiation of statin treatment, and HbA1c levels were more influential for men. The optimal risk threshold was 0.25.

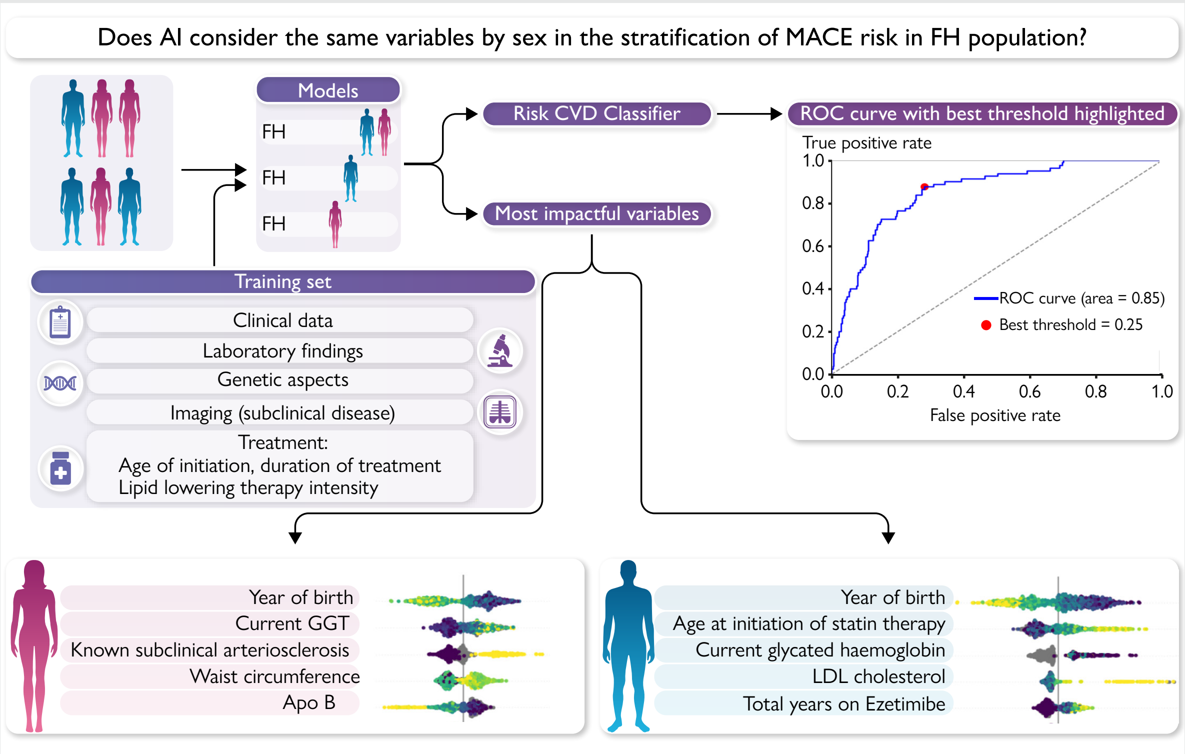
Conclusion

Artificial intelligence–machine learning algorithms are promising tools for enhancing vascular risk stratification, revealing critical sex-based differences.

* Corresponding authors. Tel: +34677852228, Email: azamora@salutms.cat; Tel: +34 645200150, Email: mcamachoruiz@gmail.com; Tel: +34610429554, Email: luis.masana@urv.cat

The institution where the work was performed: Corporació de Salut del Maresme i la Selva

Graphical Abstract



Keywords Familial hypercholesterolaemia • Artificial intelligence • Cardiovascular risk • Sex-based stratification

Introduction

Familial hypercholesterolaemia (FH) is a genetic disorder resulting in very high LDL-cholesterol (LDL-C) and increased risk for cardiovascular disease (CVD). It is the most common autosomal dominant disease, with a prevalence of around 1 in 200–250 in our environment. Currently, it is estimated that there are around 34 million patients suffering from FH worldwide. In addition, around 20–25% of the diagnosed individuals are children and teenagers.[1] It represents the most common genetic cause of early and aggressive coronary artery disease (CAD).

Without treatment, 50% of men under 50 years and 30% of women under 60 years with FH will primarily develop CAD. The life expectancy of individuals with FH has been calculated to be between 10 and 30 years lower compared with the non-FH population.[2] FH patients have a 45% higher coronary mortality after myocardial infarction and their risk of recurrence is 2.5 times higher than the general population.[3] Therefore, early identification is important to optimize treatment and to reduce extreme cardiovascular risk. Unfortunately, FH is frequently underdiagnosed and often undertreated. However, although not perfect, there are effective gold standards for FH diagnosis. Genotyping and some well-known clinical scoring tools have proven their high sensitivity and specificity. Nevertheless, new digital tools are needed to improve risk stratification in patients with HF, who are inherently at high cardiovascular risk, with particular attention to sex- and gender-specific perspectives. Recently, both the World Heart Federation and the International FH Foundation have alerted that FH is a public health priority that requires global action to improve its diagnosis and treatment and to reduce its impact on CVD.[4] They recommend the development of new screening systems such as FIND FH machine learning (ML) model[5] and the use of digital tools to improve patient risk stratification.

Risk assessment represents the first critical step in the current approach to primary prevention of atherosclerotic cardiovascular disease (ASCVD). Risk calculators cannot be used interchangeably as they have been shown to over- or underestimate cardiovascular risk in populations other than those from which they are derived.[6] The cardiovascular risk scales commonly used in clinical practice, such as Systematic Coronary Risk Evaluation Score (SCORE),[7] SCORE2,[8] and SCORE2-O,P[9] underestimate the risk in FH population. In these scales, FH patients are classified as high cardiovascular risk, and if they have other cardiovascular risk factors or have already experienced a cardiovascular event, they are considered very high cardiovascular risk.

Currently, there are only three specific risk calculators for FH: Montreal FH score designed in a Canadian population,[10,11] SAFEHEART-RE,[12] and SIDIAP-FHP.[13]

The first two are based on a genetically defined population with FH and SIDIAP-FH in patients with phenotypic FH.

While men and women share many traditional risk factors for CVD, additional gender-specific risk factors and mechanisms are at play. Therefore, it is crucial to consider gender differences when it comes to predicting and managing CVD risks. Men and women in the general[14] population and in those with FH[15] differed in the impact of the individual risk factors on the development of ASCVD. Different studies have shown that women with FH are undertreated compared to men with the same risk ASCVD, even in secondary prevention probably due to the underestimation of cardiovascular risk in women.[16,17] None of the three FH-specific risk calculators are developed specifically from a sex–gender perspective.[10–13]

ML and artificial intelligence (AI) offer promising alternatives by integrating complex datasets and providing more personalized risk assessments. Recently, ML models have been widely used to precisely predict CVD risk factors and providing a new instrument to improve early identification high risk patient, determine a patient's CVD prognosis, make better decisions in clinical practice and determinate a personalized treatment strategy.[18] In the field of FH, the use of ML and AI techniques is seen as a significant advancement for improving screening, diagnosis, and risk assessment based on various data sources, such as electronic health records (EHRs), plasma lipid profiles, genetic studies, radiology images, and corneal arcus images.[19] Therefore, it is necessary to develop robust, explainable, reliable, and ethical AI algorithms. It is crucial that these new tools incorporate the sex–gender perspective and social determinants to avoid potential biases and to prevent the continuation of a lack of information on male–female differences in the new era of digital medicine.[20]

The aim of this study is to develop an AI–ML algorithm from a sex–gender perspective useful for cardiovascular risk stratification in FH population and with a significant emphasis on model of explainable AI (XAI) to provide maximum confidence to professionals and users. For the development of the model, data from family history, clinical data, analytical data, genetic data, imaging data, and age at initiation of statin treatment, and duration and intensity of lipid-lowering treatment from the Spanish Society of Arteriosclerosis (SEA) registry were included.

Methods

Study design and population

In the present study, patients with diagnosis of FH, either with a positive genetic study or a Dutch Lipid Clinic Network score (DLCN)[21] ≥8, were selected from the National Registry of the SEA.

These patients were included in the registry between January 2013 and December 2017. It is an online, retrospective, and prospective database where accredited Spanish lipid units, recognized by the SEA, enter data from patients with lipid metabolism disorders.

Established in 2013, the registry collected clinical, analytical, genetic, and follow-up data from 4449 patients by 2017. The registry records a large volume of information in real time, including sociodemographic data, family and personal medical history, analytical results with and without treatment, genetic data, and information on lipid-lowering and other treatments. These data, gathered under strict quality criteria from 60 lipid units across all 18 regions of Spain, are entered by qualified clinical professionals.[22]

Variables

The following baseline variables were extracted from the RIHAD database:

(1) Personal and first-degree family history, sex, age, paternal hypercholesterolaemia, maternal hypercholesterolaemia, family history of CVD, age at first cardiovascular event in relatives, ischaemic heart disease, myocardial infarction, acute coronary syndrome, stable angina, coronary bypass, angioplasty, stroke, stroke type, peripheral artery disease, aortic/abdominal aneurysm, other cardiovascular events, aortic stenosis, hypertension, age at hypertension diagnosis, diabetes, smoker, non-smoker, former smoker, hepatic steatosis, and packs/day per years of smoking.
(2) Physical examination body mass index (BMI), waist, pulse, systolic blood pressure, diastolic blood pressure, tendinous xanthoma, and corneal arcus.
(3) Known subclinical arteriosclerosis: atherosclerotic lesions (incidentally detected by carotid echocardiography in asymptomatic persons; defined by a focal invasion of the carotid lumen ≥ 1.5 mm).
(4) Total DLCN score.
(5) Genetic data: LDL receptor, apoprotein B (Apo B) receptor, and PCSK9.
(6) Data on treatment and follow-up in the lipid unit. Age at statin initiation, lipid-lowering treatment, treatment duration, statin dose, total years on statins, total years on ezetimibe, and age at first visit at Lipid Unit.
(7) Laboratory values (note: for each variable, two values are provided – the baseline value at the time of admission to the Lipid Unit (first value), and the most recent value after optimization of lipid-lowering therapy (current value).
(8) Major ischaemic events: refers to the documented occurrence of major ischaemic cardiovascular events, including myocardial infarction, ischaemic stroke, and coronary revascularization procedures, as identified through clinical diagnoses recorded in the patients' medical records. Presence of at least one of the following diagnoses: CVD, CAD, ischaemic heart disease, myocardial infarction, acute coronary syndrome, stable angina, coronary bypass, angioplasty, stroke, stroke type, peripheral artery disease, or aortic abdominal aneurysm. The clinical diagnoses were extracted using the codes according to the International Classification of Diseases (ICD-10) from hospital discharge reports.

Machine learning–based approach

We followed a model based on Cross-Industry Standard Process for Data Mining (CRISP-DM).[23] The model has six steps: problem understanding, data understanding, data preparation, modelling, quality assurance, and explicability.

All AI–ML analyses and statistical analyses were conducted using Python programming language version 3.10.15 and its standard library.[24]

Problem understanding is the first step that is to obtain a model that can assist to predict the risk of a major adverse cardiovascular event (MACE) in a patient with FH. Our system must be able to take information about the patient and return a MACE risk metric that ultimately classifies the patient as either low risk or high risk. Data understanding is divided into four tasks: requirement intake, data acquisition, data exploration and quality assurance. Data preparation includes cleaning, integration, feature engineering, and scale. In our case, we discarded physical exercise frequency due to significant missing data. We apply the Standard Scale from sklearn library.

This scaler applies the following transformation:

$$z = \frac{(x–u)}{s}$$

where x is our sample, u is the mean of the training samples, and s is the standard deviation of the training samples.

Modelling: We proposed the Histogram-based Gradient Boosting Classification Tree (HGBCT) for this problem. An HGBCT is a ML algorithm used for classification tasks. It combines the principles of decision trees and boosting with the efficiency gained from binning continuous features into histograms, thus significantly improving performance on large datasets. The algorithm uses an ensemble of decision trees, where each tree corrects the errors of the previous ones. Boosting works by sequentially adding models (trees), with each new model focused on correcting the errors made by the previous ones. This helps in reducing bias and improving the model's prediction accuracy.[25] One of the main reasons to use it is that the estimator has native support for missing values. Once we had our model, we configured it with the class weight 'balanced' parameter. This means that the model uses the values of the target variable to automatically adjust weights inversely proportional to class frequencies in the input data. Data are then split into five folders following the cross-validation principle with 20/80 test/train set to train the model with a grid search for hyperparameter optimization. On top of this, we make a scorer based on F1-score and pass it to the model. This way we optimize the F1-score instead of accuracy leading to better models when unbalanced classes are present.

Once our model is configured, we train the model with 70% of data and save the remaining 30% for evaluation purposes. We ran a K-fold Cross Evaluation[26] with the base model to study its stability for this problem. It helps ensure that the mode generalizes well to unseen data by using different portions of the dataset for training and testing in multiple iterations. K-fold cross-validation was employed to ensure model stability by partitioning the dataset into k subsets, training on $k − 1$, and testing on the remaining set. We ran the model five times with different data combinations in the train and test set.

The evaluation metrics used for assessing the performance of the ML models included the following:

(1) Accuracy: This metric represents the proportion of correct predictions (both true positives [TP] and true negatives [TN]) out of the total number of predictions made by the model. It provides a general sense of the model's performance but can be misleading in imbalanced datasets, where one class is significantly more prevalent than the others.
(2) Precision: Also known as positive predictive value, precision is the proportion of TPs out of all predicted positives (i.e. the fraction of correct positive predictions). High precision indicates a low false positive (FP) rate, meaning the model is good at avoiding incorrect positive predictions.
(3) Recall: This metric reflects the ability of the model to identify TPs, defined as the proportion of actual positives that the model correctly identifies. High recall indicates that the model is capturing a large percentage of the actual positive cases, though it may also produce FPs.
(4) F1-score: It is the harmonic mean of precision and recall, providing a balanced measure when there is an uneven class distribution. It is particularly useful when both precision and recall are important, as it balances the trade-off between these two metrics.

These metrics provide a comprehensive evaluation of the ML models' performance, particularly in the context of MACE in FH patients, where both FPs and false negatives (FN) must be carefully managed.

Complementary analysis of explicability by Shapley additive explanations

It is important to explain how the model reaches a solution and the relationship of the different features with the output variable. One of the best algorithms to study explainability is Shapley Additive exPlanations (SHAP). SHAP analyse the contribution of each feature to the target variable per sample. In other words, we can estimate how any piece of data in our dataset affected the output given by the model. One way to represent this information is by plotting all samples per feature and see how they affect the outcome of the model.[27] The SHAP summary plots demonstrate the either positive or negative adjustment to MACE risk estimation (*x*-axis) for each of the predictor variables (*y*-axis). Relative values for individual predictors were represented on a continuous colour bar. A SHAP summary plot helps interpret how each variable contributes to the AI model's prediction. It shows which features have the greatest influence on the risk of cardiovascular events. Each point represents an individual patient. The horizontal position (SHAP value) shows how much that feature increases or decreases the predicted risk. Importantly, the vertical line represents the average model prediction. Points to the right of this line indicate that the feature contributed to a higher risk prediction in that patient, while points to the left indicate a lower predicted risk compared with the average. However, a lower predicted risk does not mean the patient is at *no* risk. SHAP plots help identify which features are most important and how they behave across different individuals.

Complementary analysis of Youden's *J* and its ratio

Youden's *J* and its ratio, the sum of sensitivity and specificity minus one, was used to establish the optimal cut-off point for the model associated with very high cardiovascular risk, considering that the cost is the same for an FP as for an FN. When using this index, one implicitly uses decision theory with a ratio of misclassification costs, which is equal to one minus the prevalence proportion of the disease.[28]

Ethical aspects

The present study was approved by the Ethics and Medical Research Committee of Mataró Hospital (part of the Maresme Health Consortium, Mataró, Barcelona, Spain; Code 27/24).

Results

Study population

Of the 4495 subjects in the study database at the time of inclusion, 2731 patients were excluded due to diagnoses of combined hyperlipidaemia, mixed hyperlipidaemia, polygenic hyperlipidaemia, unspecified dyslipidaemia, or FH without a positive genetic study or a DLCN score <8. Of the 1764 subjects finally included, 1540 had a positive genetic study. Among these, 95.6% had mutations in the LDLR gene, with 92.5% being simple heterozygotes, 4.1% compound heterozygotes, 2.1% double heterozygotes, and 1.2% homozygotes. The remaining subjects had a DLCN score ≥8.

Of the included patients, 127 (7%) were not from Spain. Of the total included subjects, 838 (51.8%) were women. The mean age was 50 (±26) and 48 (±15), in women and men, respectively ($P < 0.02$). A total of 264 living patients (73 women and 191 men) had a history of MACE ($P < 0.001$). Among the patients, 36.2% had a family history of premature CVD in first-degree relatives without sex differences.

Additionally, 406 patients (37%) presented with subclinical atherosclerotic disease detected by imaging techniques. Of these, 190 were women and 216 were men ($P < 0.002$).

Seventeen per cent of the subjects had a history of hypertension, and 5.8% had diabetes, with no significant differences between sexes. Twenty per cent of women and 23.8% of men were smokers ($P < 0001$). Men had a significantly higher BMI compared to women ($P < 0001$). Tendinous xanthomas were present in 39% of subjects, with no differences between sexes. The corneal arc was present in 29% of women and 39% of men ($P < 0.001$). The DLCN score was significantly lower in women (16 ± 4 vs. 17 ± 4, $P < 0.001$). The baseline LDL-C was 277 (±79.4) mg/dL with no differences by sex, while the post-treatment LDL-C was 146 (±56.5 mg/dL in women and 139 (±61.9) mg/dL in men ($P > 0.001$). The atherogenic indices Apo B/Apo A1 and triglycerides/HDL-cholesterol were significantly higher in men ($P < 0001$). The mean levels of lipoprotein(a) were 29 (11–63) mg/dL, with no significant differences between sexes.

The mean age at the first visit to the lipid unit was 42 (±15) years, with women being older than men (43 ± 16 vs. 42 ± 15 years; $P < 0.02$). Women were treated with lower doses of statins ($P < 0001$) and were less frequently prescribed combination therapy (statin + ezetimibe; $P < 0.03$). Additionally, lipid-lowering treatment was initiated later in women compared with men ($P < 0.006$).

All data presented in this section are available in Supplementary material online, *Table S1*.

Machine learning models

The model included clinical, analytical, genetic, and imaging variables as well as age at treatment initiation, intensity, and use of combination therapy (statins + ezetimibe), totalling 82 variables. The variables included in the final algorithmic model are shown in Supplementary material online, *Table S2*.

Figure 1 shows the heatmap illustrating the relationship between variables and different CVD or MACE in a population with FH.

Figure 2 presents the confusion matrices for the various ML models used to predict MACE in patients with FH. Each confusion matrix illustrates the performance of a distinct model, indicating TPs, TNs, FPs, and FNs in the prediction of MACE. The comparison includes results from five-fold cross-validation. These matrices offer a visual comparison of each model's predictive accuracy and misclassification rates, providing valuable insight into the selection of the optimal model for MACE prediction in FH patients.

The various metrics obtained from the algorithmic model for the general population and when they are divided by sex are presented in *Table 1*. Notable differences in performance metrics are observed when the algorithm, trained on the general population, is applied separately to female and male cohorts. The AI algorithm model trained for the prediction of MACE demonstrated a recall of 0.98 in the female subpopulation and 0.82 in the male subpopulation for the presence of MACE. For the absence of MACE, the recall was 0.23 for women and 0.61 for men. The F1-score for MACE was 0.96 for women and 0.85 for men, while for the absence of MACE, they were 0.32 and 0.55 for women and men, respectively. These results highlight significant

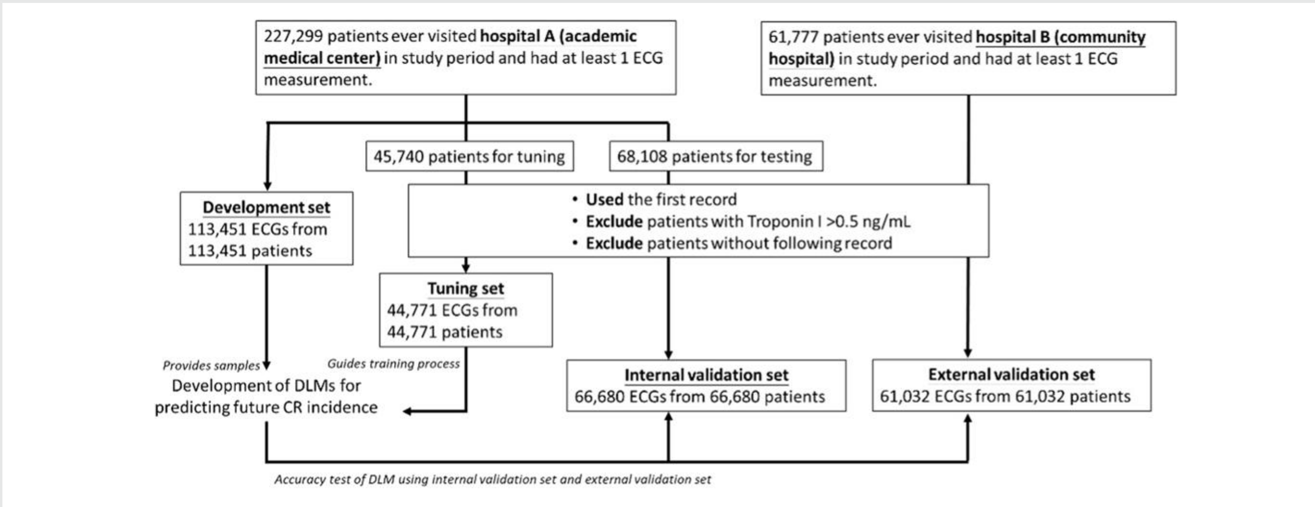


Figure 1 The heatmap displays the strength of associations between selected variables—such as lipid levels, genetic markers, and treatment history—and the occurrence of MACE. Darker colours indicate stronger correlations, either positive or negative, while lighter colours represent weaker or no associations. A threshold for statistical significance was set at $P < 0.05$. This figure provides an overview of the most impactful factors contributing to cardiovascular risk in FH patients, facilitating targeted preventive strategies and personalized management.

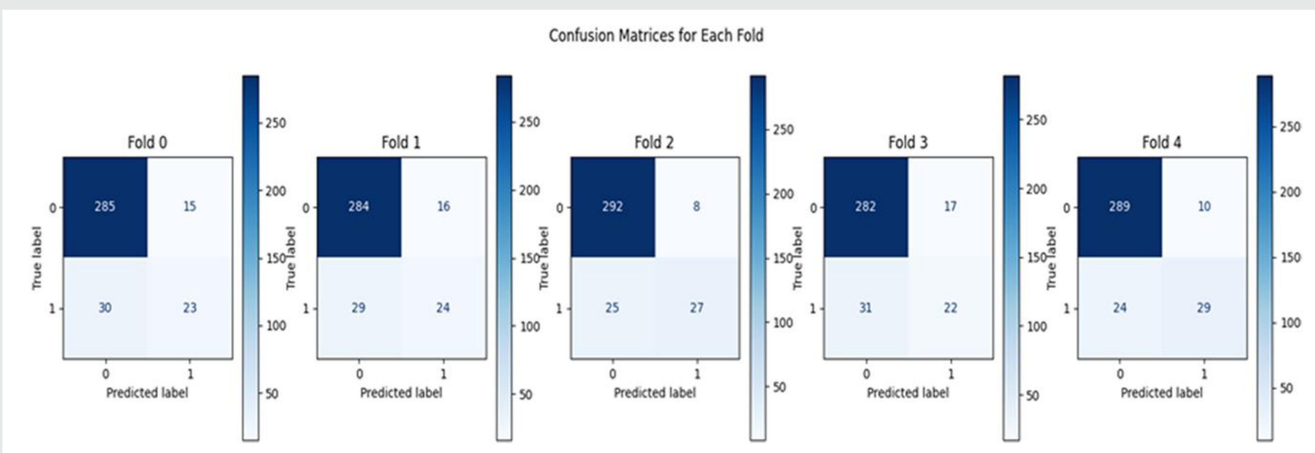


Figure 2 Each confusion matrix displays the performance of a different ML model, highlighting TP, TN, FP, and FN in the prediction of MACE. The models compared include five-folds. These matrices provide a visual comparison of each model's predictive power and potential misclassification rates, guiding the selection of the optimal model for MACE prediction in FH patients.

differences in model performance between sexes, particularly in predicting the absence of MACE.

Definition of the optimal cut-off point for the association with a pattern of high or low cardiovascular risk in FH population was assessed using Youden's index.

To define the optimal risk threshold, the population dataset used for model training was applied. Various thresholds were tested, assigning different weights to the error based on Youden's index. The optimal risk threshold, where equal weight was given to both FP and FN errors, was set at 0.25 points to define very high or extreme risk for association with MACE. The ROC curve for the AI algorithm, based on the cut-off point defined as optimal, is presented in *Figure 3A and B*. This curve illustrates the algorithm's performance at the selected threshold, highlighting the balance between sensitivity and specificity in predicting the occurrence of MACE.

Figure 4 shows the distribution of the FH population and the presence or absence of MACE based on the defined optimal risk threshold. Of the total samples, 38 exceeded this threshold; 61.4% of men and 22.7% of women with MACE were above this threshold.

Analysis of the contribution of different variables in the AI/ML model, stratified by sex, using SHAP methodology

Figures 5 and *6* display the contribution of various variables in the model created using SHAP. In women, age, current gamma-glutamyl transferase, the presence of subclinical disease, waist circumference, and Apo B were the most influential factors in the model. In men, age, age of statin initiation, current HbA1c, LDL-C, and combination therapy had the

Table 1 Performance metrics of the ML algorithm for predicting MACE and no MACE in patients with FH, presented globally and stratified by sex

Metrics	Overall	Women	Men
Accuracy MACE	0.88	0.92	0.78
Accuracy No MACE	0.84	0.92	0.78
Precision MACE	0.93	0.94	0.88
Precision No MACE	0.47	0.56	0.50
Recall MACE	0.87	0.98	0.82
Recall No MACE	0.65	0.23	0.61
F1-score MACE	0.90	0.96	0.85
F1-score No MACE	0.54	0.32	0.55

greatest impact. The importance of each variable in the model differs significantly by sex. The results highlight sex-specific differences in the importance of key variables, suggesting the need for tailored predictive models that account for these variations.

Discussion

To our knowledge, this is the first study to evaluate an AI–ML algorithm applicable to the prognostic stratification of FH population from a sex–gender perspective.

All patients with FH are considered to be at least at high cardiovascular risk; however, the degree of risk may vary, ranging from very high to extreme.[10–13] The AI–ML algorithm developed for MACE risk in the FH population in this study demonstrates the strong predictive power AI–ML: area under the curve (AUC) 0.88 [95% confidence interval (CI) 0.85–0.90], which is superior to those previously obtained with multivariable inferential models in the Spanish population (SAFEHEART-RE score: C-index 0.55 (95% CI 0.51–0.59)[12] and SIDIAP-HF score C-index: 0.71 (95% CI 0.68–0.75),[13] Indeed, it is superior to the metrics in the Canadian population (Montreal FH score: AUC of 0.79 [95% CI 0.766–0.832)].[11] The recall value in our model was moderately high (0.87 for the presence of MACE and 0.65 for the absence of MACE) but superior to the qualitative risk scales commonly used in clinical practice. In our study, 87.5% of the population with MACE, predominantly men, exceeded the threshold detected for very high cardiovascular risk. A recent analysis by the SCORE Chart showed low sensitivity in Mediterranean patients with dyslipidaemia. The scale classified 62.8% of the patients who experienced a cardiovascular event and 46.6% of those who died as low risk.[29]

Understanding the heterogeneity in risk estimation and the role of emerging biomarkers and imaging techniques is crucial for optimizing cardiovascular risk prediction and guiding personalized treatment strategies for patients with hypercholesterolaemia. A combined approach using inferential statistics and AI techniques, incorporating data from different sources, is likely a good option at this time. In a recent study by Zinzuwadia *et al.*, an ML approach enhanced the accuracy of the AHA-PREVENT model when applied to a local population while still preserving the risk associations identified by the original model. This strategy may help reclassify patients into low or high cardiovascular risk categories.[30]

The Random Survival Forest (RSF) model is the most frequently utilized model for survival outcome in CVD prediction. RSF is effective at handling complex interactions, has built in variable importance measures, and is robust to overfitting. Despite RSL models often benefiting from large datasets, they can still be effectively applied to smaller health-related datasets as long as the right balance between data quantity and quality is ensured and interpretability is prioritized.[31] Deep learning models can also be useful in cardiovascular risk stratification, but the difficulty in explaining the mode may lead to trust issues among professionals and patients.[32] In a recent observational study conducted on a population aged over 15 years in Bangladesh, the best results in predicting cardiovascular events were achieved using the RSF, with an AUC of 0.98, compared with other AI–ML models.[18] Other studies that have employed AI–ML techniques in the field of vascular risk stratification have also demonstrated improvements, particularly with the RSF, in diagnostic capability compared with inferential statistical methods, with AUCs ranging from 73 to 98.[33–38]

When the algorithm is evaluated in the subgroup divided by sex, we found different behaviours in women and men. A recent study conducted in the SAFEHEART registry using inferential statistics showed that the risk of ASCVD is markedly lower in females than in males with FH.[39] In our AI-developed model, it is observed that, generally, female sex is protective against the occurrence of cardiovascular events in the FH population. However, this protective effect disappeared, especially in younger women, when considering associations with other clinical or analytical characteristics. Age, family history of hypercholesterolaemia, aortic stenosis, the presence of subclinical disease, or waist circumference had a greater weight in the model for women, while family history of CVD, hepatic steatosis, and the presence of corneal arcus had a greater weight in men. In the resulting model, an association is observed between the presence of MACE and LDL-C, as well as the intensity of lipid-lowering treatment and the age at its initiation. However, these variables have a greater impact in men than in women. In women, Apo B levels carry more weight in the model than LDL-C levels. In a recent systematic review on the application of AI in cardiovascular risk stratification, limited evidence was found regarding sex differences.[40] Of the 31 studies that included gender in their prediction models, only 6 showed gender-stratified predictions.[41–46]

None of them were conducted in an FH population. One of these studies did not observe differences with sex and race in the discriminative power of CVD risk prediction between studies using neural networks and those conducted with pooled cohort equations when the same data were used for the analysis.[44] High-dimensional features, including diverse sources such as clinical and laboratory data, genetics, social determinants, and imaging tests and/or longitudinal risk factors, evaluating variability between visits in laboratory values and vital signs, should be considered to fully explore the benefits of neural network survival models for cardiovascular risk prediction.

It is important that ML models provide intuitive explanations that enable patients to understand their risk predictions, thereby assisting clinicians and patients in better comprehending the decision-making process for assessing disease severity and maximizing opportunities for early intervention and personalized risk prediction models. Most prior studies have focused on the performance of ML models or the importance of features, with limited attention to fully understand and explain predictions using interpretable methods, such as SHAP,[46] as in our study.[47,48]

Deep learning and neural network fields with larger datasets, related to CVD, can be explored in a future work, along with the integration of the trained ML models into proper XAI interface systems, such that the predictive results obtained from the ML models have a proper explainability, transparency, and are trusted by citizens, patients, and clinicians.[49]

The importance of this study lies in its demonstration that the use of AI–ML techniques can enhance prognostic stratification in FH populations compared with standard clinical practice. It applies a sex and gender perspective to avoid potential biases and includes explainability as a fundamental component. It is noteworthy that the data source is a national registry that includes a representative sample of patients from

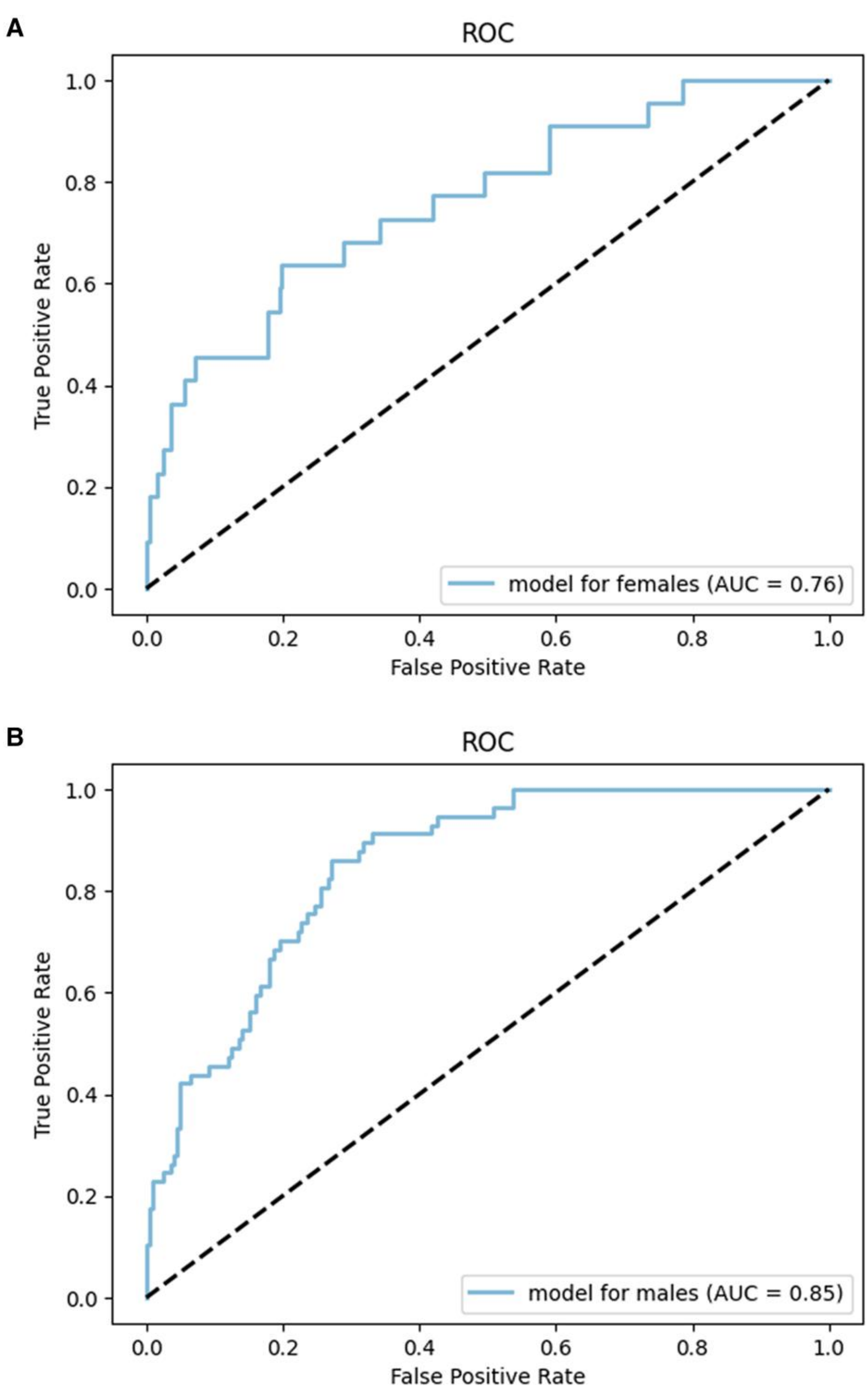


Figure 3 (*A* and *B*) The ROC curve depicts the performance of the predictive model using the selected optimal risk threshold for MACE. The curve plots sensitivity (TP rate) against 1 – specificity (FP rate) at various threshold settings. The AUC reflects the model's discriminative ability, with the optimal threshold chosen to balance sensitivity and specificity. This figure highlights the model's accuracy in identifying patients at risk of MACE within the FH population.

across the entire country with a significant number of participants. For the development of the model, different data from family history, clinical data, analytical data, genetic data, imaging data, and duration and intensity of lipid-lowering treatment were included.

This study has certain limitations. Firstly, it is a cross-sectional study that provides a snapshot of information at a specific point in time, where the output is a prevalent variable rather than an incident one. This is an initial descriptive study in which only internal validation has been

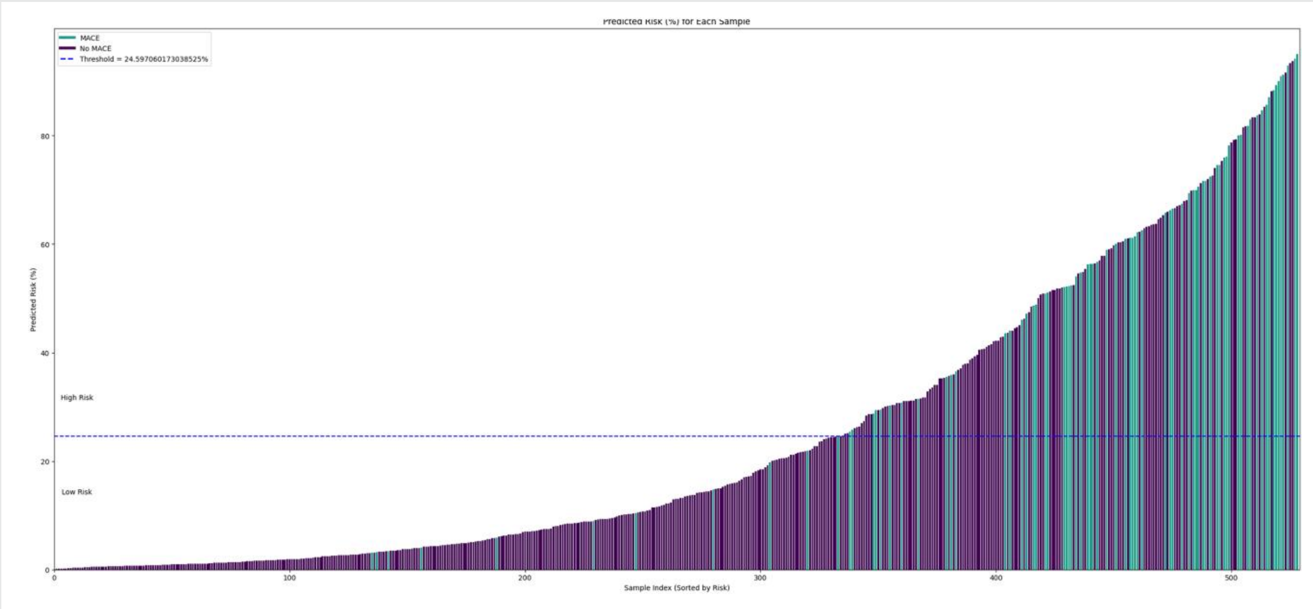


Figure 4 Each bar represents the risk of a patient and is ordered from lowest risk (left) to highest risk. The blue dashed line corresponds to the defined threshold of 0.25. Everything above it is classified as 'very high risk', while everything below is 'low risk'. The sky blue bars indicate MACE, and the violet colour bars indicate non-MACE. Sky blue colour bars below the threshold are FNs, and violet colour bars above the threshold are FPs.

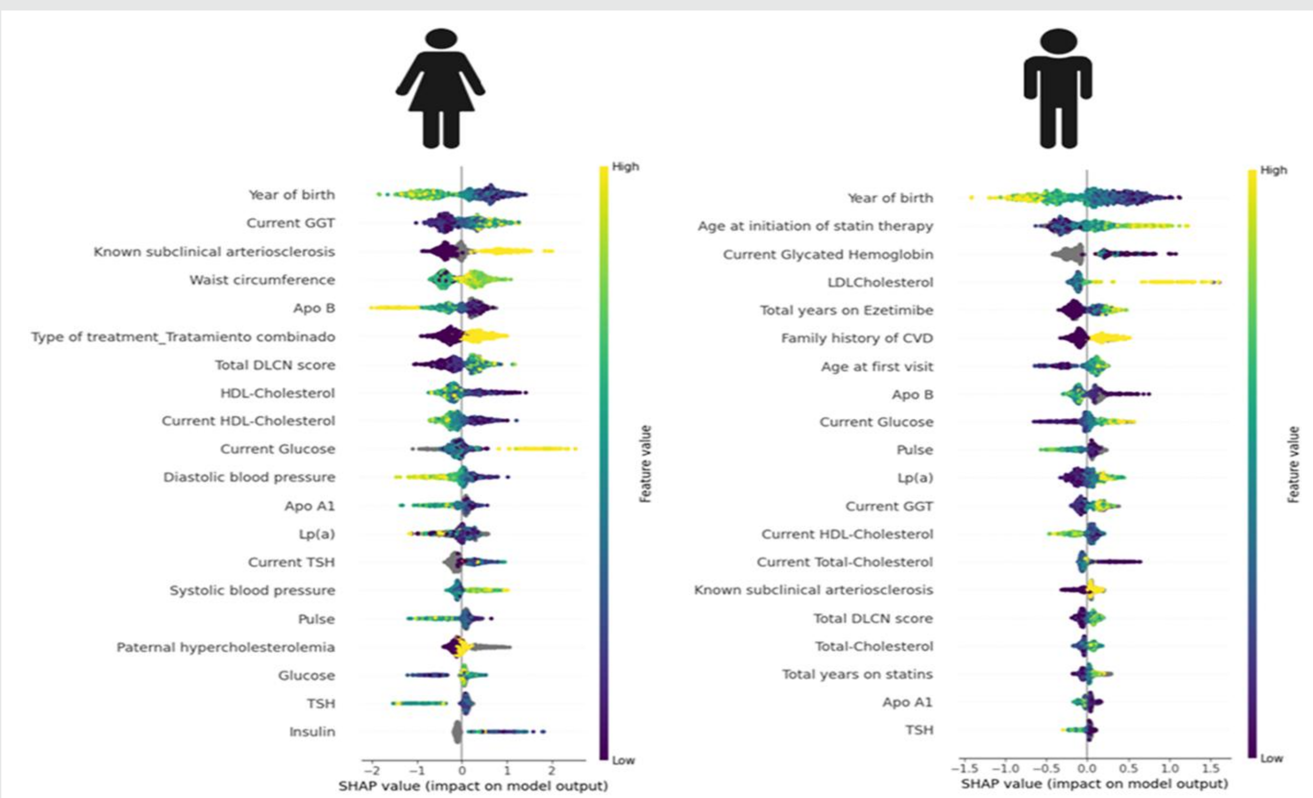


Figure 5 The coloured plot should be treated as a version of a violin plot where vertical height effects number of observations with a given SHAP value; colour of this vertical bar reflects the value of this specific variable for a given observation and horizontal distance from null reflects influence of this variable on the classification (further from the null means that this variable was more impactful on the estimated risk of this observation). Positive SHAP value means predictor variable contributed to increased risk of this observation, and negative SHAP value means predictor variable contributed to lowering the risk of a given observation.

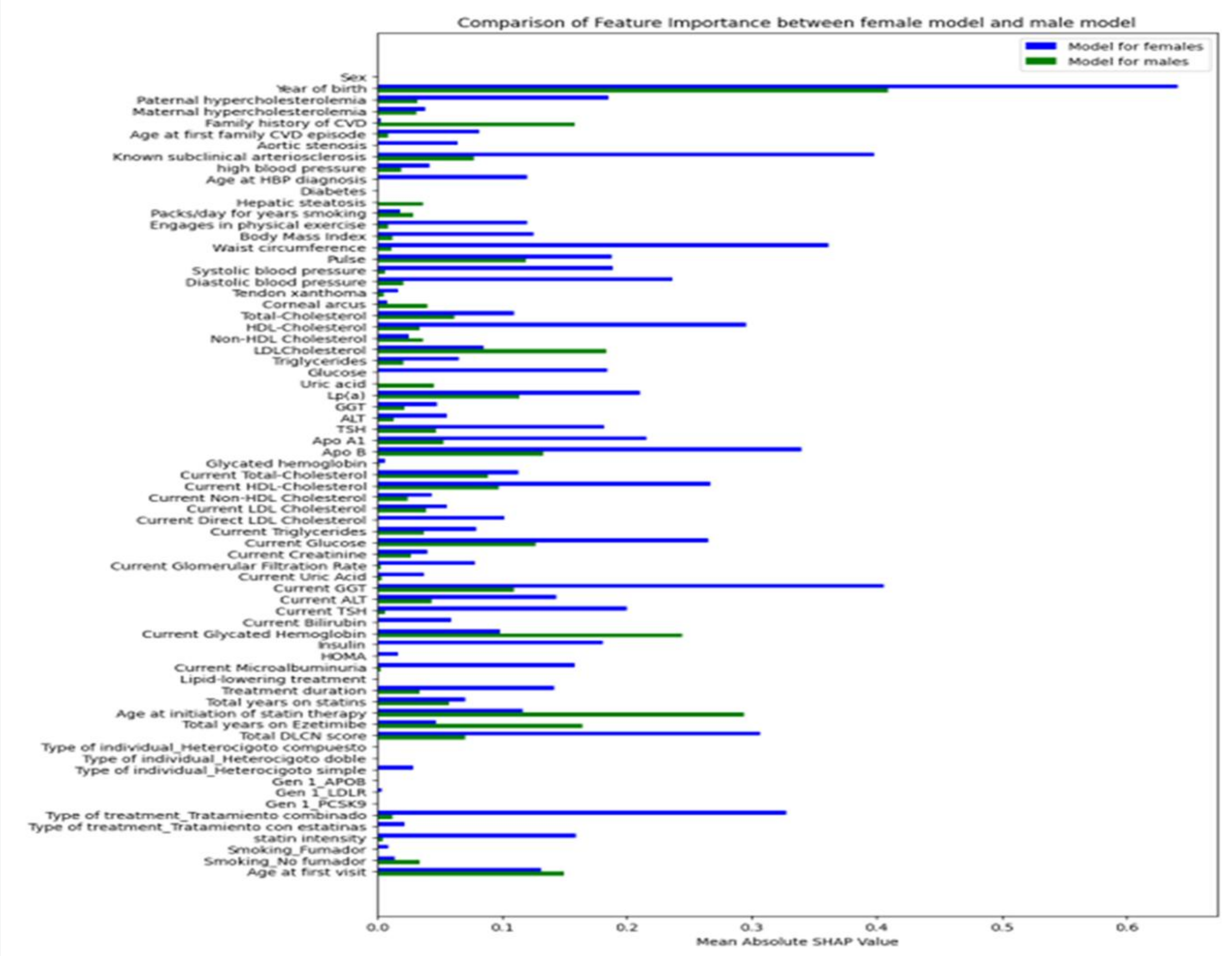


Figure 6 The SHAP plot illustrates the relative importance and impact of each variable on the model's prediction of MACE, with separate analyses for male and female patients. Each bar represents the SHAP value for a given feature, where a higher SHAP value indicates a greater contribution to MACE risk prediction. Differences between sexes are emphasized, suggesting sex-specific predictors and interactions. This figure highlights the importance of personalized risk stratification by sex in FH populations. A SHAP beeswarm plot provides a global summary of feature importance and their effect on the model's predictions across the entire dataset. (1) *Y-axis (vertical axis)* lists the features (variables) ordered from top to bottom by their overall impact on the model (most important at the top); (2) *x-axis (horizontal axis)* displays SHAP values, which quantify how much each feature contributes to increasing or decreasing the predicted risk for each individual case. Positive SHAP values push the prediction towards a higher risk (e.g. higher probability of MACE), while negative values push it towards a lower risk but no protective; (3) e*ach dot* represents a single patient. The horizontal position of the dot shows how much that specific feature influenced the prediction for that patient; and (4) *colour* reflects the actual value of the feature for that patient. Typically, red indicates a high feature value and blue indicates a low value.

performed. Any algorithm, prior to implementation, requires external validation in different populations. A second study is currently underway, in which prospective validations will be carried out in diverse populations. Secondly, it includes an imbalanced population with lower representation of individuals with MACE particularly in women. Conversely, there is likely to be an issue of collinearity among the different variables included. The robust algorithmic techniques used can minimize this effect.

Future investigations could benefit from larger sample sizes to improve the robustness of the findings, ideally using large registries such as the European FH Patient Network,[50] with particular interest in paediatric and adolescent populations, prior to widespread clinical implementation, and the perspective of sex and ethnicity is crucial to avoid future biases and ensure an ethical approach to AI.

In the future, integrating validated AI–ML algorithms into EHRs in clinical practice, using criteria for robustness, transparency, ethics, and the inclusion of social determinants, such as zip code, along with a sex–gender perspective, will enhance the management of diseases like FH, which represent a significant public health issue. Advancing towards an EHR with 'cognitive layers' powered by AI could shift medicine from a reactive to a proactive approach. Ideally, screening and stratification algorithms should be embedded into healthcare systems' electronic platforms to automatically detect citizens at very high vascular risk ('Electronic Red Flags') and prioritize preventive and health promotion strategies. Moreover, this approach opens the door to new forms of epidemiological surveillance for non-communicable diseases ('Digital Epidemiology'), potentially generating 'high vascular risk maps' to better target public health strategies. If we aim to avoid perpetuating the current sex biases in cardiovascular risk stratification in the new era of Medicine 4.0, we must prioritize the differences between men and women from the outset in the development of AI algorithms.

Conclusion

AI–ML algorithms are promising tools for enhancing vascular risk stratification in patients with FH, revealing critical sex-based differences. Further validation in larger, more diverse populations, including prospective clinical trials, is the next step before widespread clinical implementation.

Supplementary material

Supplementary material is available at European Heart Journal – Digital Health.

Acknowledgements

We would like to extend our gratitude to Dr. Adrià Pla for his valuable assistance in reviewing the manuscript for grammatical accuracy. We would like to acknowledge the use of ChatGPT 4.0 as a writing assistant in the development of this manuscript, providing support in language refinement and grammar correction. The authors, however, take full responsibility for the content and scientific integrity of the work.

Funding

With the support of Diputació de Girona, Catalonia, Spain (DipSalut) and Biomedical Research Institute of Girona Dr. Josep Trueta (IDIBGI) 1/2024 through the Grants for Research Projects from Regional Hospitals of Girona – 2025 Call The sponsors had no role in the design and conduct of the study; the collection, management, analysis, or interpretation of the data; or the preparation, review, or approval of the manuscript.

Conflict of interest: None declared.

Data availability

The Dyslipidemia Database of the Spanish Society of Atherosclerosis has been used, which integrates pseudonymized information from all Lipid Units across Spain. This database is hosted on a server that complies with healthcare standards for clinical data security and has a specific cybersecurity and contingency plan in place.

References

1. Tokgozoglu L, Kayikcioglu M. Familial hypercholesterolemia: global burden and approaches. *Curr Cardiol Rep* 2021;**23**:151.
2. Beheshti SO, Madsen CM, Varbo A, Nordestgaard BG. Worldwide prevalence of familial hypercholesterolemia: meta-analyses of 11 million subjects. *J Am Coll Cardiol* 2020;**75**: 2553–2566.
3. Pérez de Isla L, Alonso R, Mata N, Saltijeral A, Muñiz O, Rubio-Marin P, *et al.* Coronary heart disease, peripheral arterial disease, and stroke in familial hypercholesterolaemia: insights from the SAFEHEART registry (Spanish familial hypercholesterolaemia cohort study). *Arterioscler Thromb Vasc Biol* 2016;**36**:2004–2010.
4. Representatives of the Global Familial Hypercholesterolemia Community; Wilemon KA, Patel J, Aguilar-Salinas C, Ahmed CD, Alkhnifsawi M. Reducing the clinical and public health burden of familial hypercholesterolemia: a global call to action. *JAMA Cardiol* 2020; **5**:217–229.
5. Myers KD, Knowles JW, Staszak D, Shapiro MD, Howard W, Yadava M, *et al.* Precision screening for familial hypercholesterolaemia: a machine learning study applied to electronic health encounter data. *Lancet Digit Health* 2019;**1**:e393–e402.
6. Bosco G, Mszar R, Piro S, Sabouret P, Gallo A. Cardiovascular risk estimation and stratification among individuals with hypercholesterolemia. *Curr Atheroscler Rep* 2024;**26**: 537–548.
7. Mach F, Baigent C, Catapano AL, Koskinas KC, Casula M, Badimon L, *et al.* 2019 ESC/EAS Guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk. *Eur Heart J* 2020;**41**:111–188.
8. SCORE2 Working Group and ESC Cardiovascular Risk Collaboration. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. *Eur Heart J* 2021;**42**:2439–2454.
9. SCORE2-OP Working Group and ESC Cardiovascular Risk Collaboration. SCORE2-OP risk prediction algorithms: estimating incident cardiovascular event risk in older persons in four geographical risk regions. *Eur Heart J* 2021;**42**:2455–2467.
10. Paquette M, Dufour R, Baass A. The Montreal-FH-SCORE: a new score to predict cardiovascular events in familial hypercholesterolemia. *J Clin Lipidol* 2017;**11**:80–86.
11. Paquette M, Brisson D, Dufour R, Khoury É, Gaudet D, Baass A. Cardiovascular disease in familial hypercholesterolemia: validation and refinement of the Montreal-FH-SCORE. *J Clin Lipidol* 2017;**11**:1161–1167.e3.
12. Pérez de Isla L, Alonso R, Mata N, Fernández-Pérez C, Muñiz O, Díaz-Díaz JL, *et al.* Predicting cardiovascular events in familial hypercholesterolemia: the SAFEHEART registry (Spanish familial hypercholesterolemia cohort study). *Circulation* 2017;**135**: 2133–2144.
13. Ramos R, Masana L, Comas-Cufí M, García-Gil M, Martí-Lluch R, Ponjoan A, *et al.* Derivation and validation of SIDIAP-FHP score: a new risk model predicting cardiovascular disease in familial hypercholesterolemia phenotype. *Atherosclerosis* 2020;**292**: 42–51.
14. Feldman RD. Sex-specific determinants of coronary artery disease and atherosclerotic risk factors: estrogen and beyond. *Can J Cardiol* 2020;**36**:706–711.
15. Klevmoen M, Mulder JWCM, Roeters van Lennep JE, Holven KB. Sex differences in familial hypercholesterolemia. *Curr Atheroscler Rep* 2023;**25**:861–868.
16. Jiménez A, Viñals C, Marco-Benedí V, González P, Domenech M, Suárez-Tembra M, *et al.* Sex disparities in familial hypercholesterolemia. *J Am Coll Cardiol* 2023;**81**:203–205.
17. Zamora A, Ramos R, Comas-Cufi M, García-Gil M, Martí-Lluch R, Plana N, *et al.* Women with familial hypercholesterolemia phenotype are undertreated and poorly controlled compared to men. *Sci Rep* 2023;**13**:1492.
18. Hossain S, Hasan MK, Faruk MO, Aktar N, Hossain R, Hossain K. Machine learning approach for predicting cardiovascular disease in Bangladesh: evidence from a cross-sectional study in 2023. *BMC Cardiovasc Disord* 2024;**24**:214.
19. Luo RF, Wang JH, Hu LJ, Fu QA, Zhang SY, Jiang L. Applications of machine learning in familial hypercholesterolemia. *Front Cardiovasc Med* 2023;**10**:1237258.
20. Teshale AB, Htun HL, Vered M, Owen AJ, Freak-Poli R. A systematic review of artificial intelligence models for time-to-event outcome applied in cardiovascular disease risk prediction. *J Med Syst* 2020;**49**:68.
21. Defesche JC, Lansberg PJ, Umans-Eckenhausen MA, Kastelein JJ. Advanced method for the identification of patients with inherited hypercholesterolemia. *Semin Vasc Med* 2004; **4**:59–65.
22. Pérez-Calahorra S, Sánchez-Hernández RM, Plana N, Valdivielso P, Civeira F. National dyslipidemia registry of the Spanish Arteriosclerosis Society: current status. *Clin Investig Arterioscler* 2017;**29**:248–253.
23. Schröer C, Kruse F, Gómez JM. A systematic literature review on applying CRISP-DM process model. *Procedia Comput Sci* 2021;**181**:526–534.
24. Python Software Foundation. Python. n.d. https://www.python.org/ (5 October 2024).
25. Nhat-Duc H, Van-Duc T. Comparison of histogram-based gradient boosting classification machine, random forest, and deep convolutional neural network for pavement raveling severity classification. *Autom Constr* 2023;**148**:104767.
26. Dutschmann TM, Kinzel L, Ter Laak A, Baumann K. Large-scale evaluation of k-fold cross-validation ensembles for uncertainty estimation. *J Chem inform* 2023;**15**:49.
27. Slack D, Hilgard S, Jia E, Singh S, Lakkaraju H. Fooling LIME and SHAP: adversarial attacks on post hoc explanation methods. In: *AIES '20: Proceedings of the AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society. New York, NY, USA, February 7–9*, 2020. p.180–186.
28. Smits N. A note on Youden's *J* and its cost ratio. *BMC Med Res Methodol* 2010;**1**:89.
29. Bertomeu-González V, Soriano Maldonado C, Bleda-Cano J, Carrascosa-Gonzalvo S, Navarro-Perez J, López-Pineda A, *et al.* Predictive validity of the risk SCORE model in a Mediterranean population with dyslipidemia. *Atherosclerosis* 2019;**290**:80–86.
30. Zinzuwadia AN, Mineeva O, Li C, Farukhi Z, Giulianini F, Cade B, *et al.* Tailoring risk prediction models to local populations. *JAMA Cardiol* 2024;**9**:1018–1028.
31. Ali S, Abuhmed T, El-Sappagh S, Muhammad K, Alonso-Moral JM, Confalonieri R, *et al.* Explainable artificial intelligence (XAI): what we know and what is left to attain trustworthy artificial intelligence. *Inf Fusion* 2023;**99**:101805.
32. Huang Y, Li J, Li M, Aparasu RR. Application of machine learning in predicting survival outcomes involving real-world data: a scoping review. *BMC Med Res Methodol* 2023; **23**:268.
33. Hossain MI, Maruf MH, Khan MAR, Prity FS, Fatema S, Ejaz MS, *et al.* Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison. *Iran J Comput Sci. 2023* 2023;**6**:397–417.
34. Ali MM, Paul BK, Ahmed K, Bui FM, Quinn JMW, Moni MA. Heart disease prediction using supervised machine learning algorithms: performance analysis and comparison. *Comput Biol Med* 2021;**136**:104672.

35. Kumar NK, Sindhu GS, Prashanthi DK, Sulthana AS. Analysis and prediction of cardio vascular disease using machine learning classifiers. In: *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, Tamil Nadu, India, March 6–7, 2020*. p.15–21. IEEE.
36. Fahim KE, Yassin H, Amin MH, Dewan PD, Islam A. Detection of cardiovascular disease of patients at an early stage using machine learning algorithms. In: *2022 International Conference on Healthcare Engineering (ICHE), September 23, 2022, in Johor Bahru, Malaysia*. IEEE. p.1–6.
37. Amzad Hossen MD, Tazin T, Khan S, Alam E, Sojib HA, Khan MM, *et al.* Supervised machine learning-based cardiovascular disease analysis and prediction. *Math Probl Eng* 2021; **2021**:1–10.
38. Nashif S, Raihan MR, Islam MR, Imam MH. Heart disease detection by using machine learning algorithms and a real-time cardiovascular health monitoring system. *World J Eng Technol* 2018;**6**:854–873.
39. de Isla LP, Vallejo-Vaz AJ, Watts GF, Muñiz-Grijalvo O, Alonso R, Diaz-Diaz JL, *et al.* Long-term sex differences in atherosclerotic cardiovascular disease in individuals with heterozygous familial hypercholesterolaemia in Spain: a study using data from SAFEHEART, a nationwide, multicentre, prospective cohort study. *Lancet Diabetes Endocrinol* 2024;**12**:643–652.
40. Teshale AB, Htun HL, Vered M, Owen AJ, Freak-Poli R. A systematic review of artificial intelligence models for time-to-event outcome applied in cardiovascular disease risk prediction. *J Med Syst* 2024;**48**:68.
41. Barbieri S, Mehta S, Wu B, Bharat C, Poppe K, Jorm L, *et al.* Predicting cardiovascular risk from national administrative databases using a combined survival analysis and deep learning approach. *Int J Epidemiol* 2022;**51**:931–944.
42. Blanchard M, Feuilloy M, Sabil A, Gerves-Pinquie C, Gagnadoux F, Girault JM. A deep survival learning approach for cardiovascular risk estimation in patients with sleep apnea. *IEEE Access* 2022;**10**:133468–133478.
43. Chun M, Clarke R, Cairns BJ, Clifton D, Bennett D, Chen Y, *et al.* Stroke risk prediction using machine learning: a prospective cohort study of 0.5 million Chinese adults. *J Am Med Inform Assoc* 2021;**28**:1719–1727.
44. Deng Y, Liu L, Jiang H, Peng Y, Wei Y, Zhou Z, *et al.* Comparison of state-of-the-art neural network survival models with the pooled cohort equations for cardiovascular disease risk prediction. *BMC Med Res Methodol* 2023;**23**:22.
45. Qian X, Keerman M, Zhang X, Guo H, He J, Maimaitijiang R, *et al.* Study on the prediction model of atherosclerotic cardiovascular disease in the rural Xinjiang population based on survival analysis. *BMC Public Health* 2023;**23**:1041.
46. Salah H, Srinivas S. Explainable machine learning framework for predicting long-cardiovascular disease risk among adolescents. *Sci Rep* 2022;**12**:21905.
47. Sharma S, Parmar M. Heart diseases prediction using deep learning neural network model. *Int J Innov Technol Explor Eng* 2020;**9**:124–137.
48. Miranda E, Adiarto S, Bhatti FM, Zakiyyah AY, Aryuni M, Bernando C. Understanding arteriosclerotic heart disease patients using electronic health records: a machine learning and Shapley Additive exPlanations approach. *Healthc Inform Res* 2023;**29**:228–238.
49. Guleria P, Naga Srinivasu P, Ahmed S, Almusallam N, Alarfaj FK. XAI framework for cardiovascular disease prediction using classification techniques. *Electronics (Basel)* 2022; **11**:4086.
50. Bedlington N, Abifadel M, Beger B, Bourbon M, Bueno H, Ceska R, *et al.* The time is now: achieving FH paediatric screening across Europe—the Prague declaration. *GMS Health Innov Technol* 2022;**16**:Doc04.