medRxiv preprint doi: https://doi.org/10.1101/2023.08.21.23293418; this version posted November 12, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Iron biomarkers predict peripheral artery disease in females: a cross-sectional analysis of HEIST-DiC and NHANES participants

Anand Ruban Agarvas[1,2], Stefan Kopf[1,3,4], Tiago J.S. Lopes[1,2,10], Paul Thalmann[1,5], José Manuel Fernández-Real[6], Peter Nawroth[1,3], *Martina U. Muckenthaler[1,2,7,8,9]

Affiliations

[1]Heidelberg University, Heidelberg, Germany.

[2]Center For Translational Biomedical Iron Research, Department of Pediatric Hematology, Oncology Immunology and Pulmonology, University Hospital Heidelberg, Heidelberg, Germany.

[3]Clinic for Endocrinology, Diabetology, Metabolic Diseases and Clinical Chemistry (Internal Medicine 1), University Hospital Heidelberg, Heidelberg, Germany.

[4]German Center of Diabetes Research (DZD), Neuherberg, Germany.

[5]Institute of Medical Biometry, University Hospital Heidelberg, Heidelberg, Germany.

[6]Department of Diabetes, Endocrinology and Nutrition, Institut d'Investigació Biomèdica de Girona (IdIBGi), CIBEROBN (CB06/03/010) and Instituto de Salud Carlos III (ISCIII), Girona 17007, Spain.

[7]Molecular Medicine Partnership Unit, Heidelberg, Germany.

[8]Translational Lung Research Center Heidelberg (TLRC), German Center for Lung Research (DZL), Heidelberg, Germany.

[9]German Centre for Cardiovascular Research (DZHK), Partner Site Heidelberg/Mannheim, Germany.

[10]Nezu Life Sciences, Karlsruhe, Germany

**Corresponding author*

Address for correspondence

Department of Pediatric Hematology, Oncology and Immunology, Heidelberg University, Heidelberg, Germany.
Tel.: +496221566923; Fax: +496221564580. E-mail: martina.muckenthaler@med.uni-heidelberg.de (M.U. Muckenthaler).

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.08.21.23293418; this version posted November 12, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

Background

Iron overload has been proposed as a risk factor for atherosclerosis but the available data are controversial. Here we investigated whether iron biomarkers within physiological limits that frequently differ between sexes are associated with peripheral arterial disease (PAD)

Methods

Using two different analytical approaches (machine-learning and logistic regression), we studied the association between blood iron biomarkers and PAD in 368 individuals from the Heidelberg Study on Diabetes and Complications (HEIST-DiC) and in 5101 individuals from the National Health and Nutrition Examination Survey (NHANES 1999-2004).

Results

We found that iron biomarkers were among the top predictors of PAD in the machine-learning classification in both cohorts. In the HEIST-DiC cohort, ferritin, iron, and transferrin were ranked among the top predictive markers, while in the NHANES cohort ferritin, total iron binding capacity (TIBC), transferrin saturation (TSAT) and iron showed high predictive power. In the regression analysis ferritin showed a positive interaction among females in the HEIST-DiC (OR 2.68, 95% CI 0.94-7.61, *p*=0.057) and NHANES cohorts (OR 1.76, 95% CI 1.16-2.67, *p*=0.008). The multivariable regression analysis of the NHANES cohort, detected a nonlinear relationship between ferritin and PAD in that certain ferritin ranges (48-97 ng/mL: OR 14.59, 95% CI 1.6-135.93, p= 0.019; 98-169 ng/mL: OR 171.07, 95% CI 1.27-23404, p= 0.039) in females associated positively with PAD.

Conclusion

Taken together our data show that iron biomarkers, importantly, elevated ferritin is associated with clinically apparent PAD in females. These findings add to the body of evidence suggesting sex differences in PAD and highlight a possible role of iron (directly or indirectly) in this relationship.

Keywords

Iron, Vascular Diseases, Arterial Occlusive Diseases, Atherosclerosis, Peripheral Arterial Disease

medRxiv preprint doi: https://doi.org/10.1101/2023.08.21.23293418; this version posted November 12, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

Peripheral arterial disease (PAD) is a common atherosclerotic vascular disorder of the arteries of the lower limbs that affects millions of individuals worldwide[1,2]. Compared to cardiovascular and cerebrovascular disease, PAD is less well studied, and our understanding of the disease process is still poor. The known risk factors of PAD include old age, smoking, and comorbidities such as diabetes, hypertension or Vitamin D excess[1].

Iron is an essential micronutrient, however, its accumulation in excess contributes to the generation of reactive oxygen species. Various regulatory systems maintain iron homeostasis by controlling absorption, transport, and storage[3]. Under physiological conditions, iron is bound to transferrin in the plasma, and transported to cells that require iron. Usually, less than one-third of transferrin's binding sites are occupied by iron, allowing free iron entering the circulation to be efficiently sequestered. However, in pathological states associated with iron overload, the binding capacity of transferrin is nearly saturated resulting in the appearance of 'non-transferrin-bound iron (NTBI)'. Therefore, iron overload is discussed as a risk factor for the development of atherosclerosis[4]. Accordingly, inhibition of ferroptosis, an iron-dependent form of cell death, has been shown to protect from atherosclerosis[5]. Iron and PAD may be linked by sex hormones. Changes in estrogen and progesterone levels associated with menopause were shown to impair endothelial function, especially among women[6,7]. Both hormones further control hepcidin, the master regulator of iron homeostasis in opposing directions, with estrogen[8] inhibiting and progesterone[9] activating hepcidin expression.

In our previous work, we demonstrated that iron overload in mouse models and humans with hemochromatosis show an aggravation of atherosclerosis and vessel damage, respectively[10]. Specifically, we found that NTBI in the circulation contributes to atherosclerosis by oxidising low-density lipoproteins (LDL) and by inducing vascular dysfunction; in addition, iron deposition in the vasculature of hemochromatotic mice was associated with more severe atherosclerosis[10]. Whether or not iron levels within physiological limits affect atherogenesis is currently unclear and available data are controversial. Most studies undertaken so far have investigated the relationships between iron-related parameters [e.g. plasma iron, ferritin, transferrin, transferrin saturation (TSAT)] and cardiovascular disease[11–17] but associations with PAD are rarely addressed.

To address this, we analyzed iron biomarkers for their associations with PAD and by considering sex-specific differences. We used two different datasets and orthogonal approaches: 1) machine learning (ML) classification and 2) logistic regression. Both methods

medRxiv preprint doi: https://doi.org/10.1101/2023.08.21.23293418; this version posted November 12, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

can be used for analyzing relationships between variables, but they differ in their approach and application. ML classification tasks can analyze complex patterns in multidimensional data and identify nonlinear interactions among variables. They are particularly useful when manual specification of relationships is challenging as is the case for human clinical data. In contrast, regression analysis quantifies the strength and direction of associations between variables and provides interpretable results in terms of odds ratios. Both approaches carry their strengths and limitations. However, we expect that true associations in the data (in this case, sex-specific differential associations between iron biomarkers and PAD) should be detected by both methods. Therefore, we initially used the ML approach to provide unbiased insight into our data and subsequently regression analyses to validate identified relationship. An outline of the study is shown in Figure 1.

Methods

HEIST-DiC study

The Heidelberg Study on Diabetes and Complications (HEIST-DiC; https://clinicaltrials.gov; NCT03022721) is a prospective, observational study designed to analyze the development of complications in individuals with diabetes and prediabetes. We analyzed the data from 368 individuals enrolled in the HEIST-DiC study. The selection criteria for participants have been described previously[18]. We excluded participants with values of serum iron biomarkers suggestive of hemochromatosis[19]. In addition, since ferritin not only indicates iron stores, but also is an acute phase protein that responds to inflammation, we excluded participants with ferritin> 400 ng/mL. For all participants, a detailed medical history was obtained. The ankle-brachial index (ABI) was measured using noninvasive blood pressure measurements of arms and ankles (ABI System 1000; Boso d.o.o.). Blood samples were obtained in the fasting state and all baseline parameters were analyzed under standardized conditions in the central laboratory of the University Hospital of Heidelberg. Serum samples were stored at -80C until further analysis. We analyzed iron, ferritin, and TSAT as biomarkers of systemic iron status, and assayed serum intercellular adhesion molecule-1 (ICAM1) and vascular-cell adhesion molecule-1 (VCAM1) as markers of vascular dysfunction. We measured ICAM1 (#171B6009M) and VCAM1 (#171B6022M) in serum using multiplex magnetic bead-array-based technology on a BioPlex200 system (Bio-Rad) according to the manufacturer's instructions. All participants provided written informed consent, and the ethics committees of the University of Heidelberg approved this study (Decision No. 204/2004, 400/2010, and S-383/2016) according to the Declaration of Helsinki.

NHANES

medRxiv preprint doi: https://doi.org/10.1101/2023.08.21.23293418; this version posted November 12, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

The National Health and Nutrition Examination Survey (NHANES) is designed to assess the health and nutritional status of adults and children in the United States. The survey includes both an interview (in which participants are asked questions about their health and lifestyle), and a physical examination component (in which various measurements and samples are collected). We obtained data from 15,969 participants over 40 years in 3 consecutive cycles of NHANES (1999-2004) for whom ABI was measured.

The demographic variables, age (in years) and body mass index (BMI) were used as is. Smoking and alcohol were categorized according to a previous study[20]. We further categorized diabetes based on self-reported history, current use of insulin or oral hypoglycemic agents, fasting plasma glucose (FPG) ≥7.0 mM, or glycated hemoglobin (HbA1c) ≥6.5%. Prediabetes was defined among those with an absent self-reported history of diabetes by either FPG 5.6-6.9 mM or HbA1c 5.7-6.4%. We categorized hypertension based on self-reported history or current use of anti-hypertensive agents or systolic blood pressure >130 mmHg or diastolic blood pressure >80 mmHg. The presence of nonalcoholic fatty liver disease (NAFLD) was categorized by calculating the Fibrosis-4 Index score (FIB-4 score>3.25)[21]. As an index of insulin resistance, we calculated HOMA2-IR using C-peptide values instead of insulin[22]. The use of antiplatelet and antihyperlipidemic medications was categorized based on medication intake.

We applied the following exclusion criteria: not even one ABI measurement available, likely hemochromatosis[19], anemia (hemoglobin <10g/L), elevated ferritin (>400 ng/mL), cancer, on current hormonal medications or iron supplements, donated blood in the last 1 month, with features of kidney disease (on dialysis or severe albuminuria), pregnant or breastfeeding in the last year, high risk of liver fibrosis, concurrent inflammation [C-Reactive Protein (CRP) >3 mg/dL] and with a feature of familial hyperlipidemia [LDL-Cholesterol (LDLc) >4.9 mM]. This resulted in a final sample number of 5101 participants. We obtained data from the following iron biomarkers: iron, ferritin, total iron binding capacity (TIBC), and TSAT. Assays for the analytes in NHANES were performed under strict laboratory protocols (for more details on data fields and procedure manuals, see Supplementary Table 1).

Categorization as PAD

In both cohorts, ABI values were measured on both the right and left limbs. We categorized PAD when the ABI value on one of the sides was <0.9 or >1.4 as per recommendations[23].

Implementation of the ML framework

medRxiv preprint doi: https://doi.org/10.1101/2023.08.21.23293418; this version posted November 12, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

First, we preprocessed the HEIST-DiC and the NHANES dataset and removed all features that contained more than 10% of missing values. Additionally, to made both cohorts comparable by only considering the cases that did not have missing values for ferritin, diabetes, CRP and high blood pressure in the NHANES cohort. For the HEIST-DiC dataset, we have used the leave-one-out split strategy, and for each fold, we filled in the missing values using the average of each column (SimpleImputer method, from Scikit-learn). Moreover, we used SMOTE (v0.12.0)[24] to oversample the minority class (PAD), to match the majority class (No PAD). For the NHANES dataset, we used the 10-fold cross-validation strategy, and for each fold, we filled the missing values and oversampled the minority class using the same methods, and performed the so-called one-hot encoding, where we encoded the categorical variables into numerical values. Next, for each dataset separately, we trained the XGBoost classifier (v2.0.3)[25] and predicted the class of the instances on the test set, from where we derived the balanced accuracy, the F1 Score, and quantified the feature importance. Finally, we also implemented a version of the ML framework that allowed us to assess the variable importance using SHAP (SHapley Additive exPlanations, v0.44.0).[26] Here, we performed the same steps for both HEIST and NHANES datasets, except that we did not use any split strategy; instead, we filled the missing values, oversampled the minority class and trained the classifier using the complete dataset. The ML pipeline is shown in Figure 2A.

Regression analysis

For the regression analysis, we first tested the interaction effect of each of the iron biomarkers with sex in both cohorts as iron metabolism is hallmarked by sex-specific differences. Further, in the NHANES cohort, we conducted multivariable regression analyses by flexibly modelling the probability of PAD's presence with one of the iron biomarkers (iron, ferritin, or TSAT) by spline regression (degree of freedom=4). We chose variables from the feature importance in ML analysis as covariates in the regression models on the NHANES data: age, CRP, HDLc, triacylglycerols, HbA1c, ACR, BMI, ethnicity (non-Hispanic whites as the reference), smoking (non-smokers as reference), the presence of diabetes, AST, ALT, Hb, creatinine, antihyperlipidemic medications; the former seven variables as continuous and the rest as categorical). All continuous variables were mean-centered for the analysis, and variables were included in the models only when their missing data was <10%. In addition, sample weights were included in the analyses of NHANES data.

All regression analyses were performed using R Statistical Software (v4.2.2[27]). We used Python (v3.10.12[28]), and Scikit-learn (v1.4.0[29]) for the implementation of the ML framework. The reporting of the study is as per The Strengthening the Reporting of Observational

medRxiv preprint doi: https://doi.org/10.1101/2023.08.21.23293418; this version posted November 12, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Studies in Epidemiology (STROBE) guidelines[30]. We have also developed a data exploration tool for readers to explore the datasets and interactively create their visualizations from the paper (https://anandr.shinyapps.io/NHANES_HEISTDiC/) using Quarto[31] and Shiny[32].

Results

Demographics

The baseline characteristics of the HEIST-DiC participants are shown in Table 1. The overall prevalence of PAD was 6.6% with a higher proportion among males (17/192) than females (6/165). Among both sexes, the participants with PAD were older and diabetes and hypertension was a common feature (Table 1). The sICAM1 and sVCAM1 levels were significantly higher only among males with PAD (Table 1). Among the iron biomarkers, ferritin was elevated among females, but not males, with PAD (Median 141, IQR 93-254, $p=0.035$) while iron, transferrin, and TSAT were similar between HEIST-DIC participants with and without PAD. (Table 1).

The analysis of NHANES builds on a previous report[33] by including data from the NHANES 2003-2004 cycle. The baseline characteristics of the NHANES participants included in the study are shown in Table 2. Overall, the prevalence of PAD in the population was 6.25% with a significantly higher prevalence among females (males: 4.96%, females: 7.9%; $p=0.001$). Individuals with PAD showed differences in ethnicity, sex, presence of diabetes, smoking, alcohol status, physical activity, and use of antihyperlipidemic and antiplatelet medications (Table 2). Among the iron biomarkers, serum iron (Median 80, IQR 64-108 µg/dL; $p<0.001$) and transferrin saturation (Median 24, IQR 17-30; $p=0.021$) were reduced in males with PAD while females with PAD showed higher ferritin (Median 87, IQR 56-252 ng/mL; $p<0.001$) and lower TIBC (Median 60, IQR 54-68; $p<0.001$). Differences were also observed in some laboratory parameters within both sexes with PAD (e.g. albumin, AST, urea, creatinine, CRP, and ACR; Table 2).

Establishing a machine learning predictive framework for PAD

We next aimed to establish a predictive ML framework to anticipate the occurrence of PAD based on biomarkers. A so-called ML classifier is trained on a dataset containing instances with known outcomes (i.e., the presence or absence of PAD), and its relationship with input features. During the test phase, the trained classifier is used to predict PAD in participants unseen during the training phase, where we purposefully hid the true diagnostic. The result is then compared to the real diagnosis, and the quality of the framework is evaluated. Overall, this process enables the identification of subtle correlations between variables that

medRxiv preprint doi: https://doi.org/10.1101/2023.08.21.23293418; this version posted November 12, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

might be imperceptible to human analysis, but that can be used to anticipate PAD and enhance our diagnostic capabilities.

Our framework received the available molecular and physiological biomarkers from participants from two distinct cohorts, HEIST-DiC and NHANES as input. These cohorts have unique characteristics that require the development of separate data preprocessing, training, and testing regimens. First, the HEIST-DiC cohort comprised 357 instances (23 PAD and 334 No PAD cases) and had 35 numerical and categorical features. Here, for the training and testing of the ML framework, we used a leave-one-out split methodology. This approach involves using each instance as a test case once, while the remainder forms the training set, thus maximizing the use of available data for training. In contrast, the NHANES cohort was considerably larger, with 1188 instances (100 PAD and 1088 No PAD), and 28 features; hence, we used the 10-fold cross-validation strategy. This method divides the dataset into ten parts, with each part successively used for testing while the other 9 parts are used for training. Together, these strategies ensured robustness and reliability in our model's predictive performance.

We found that the ML framework achieved a balanced accuracy of 95% for the HEIST cohort (Figure 2B-D) and 96% for the NHANES cohort (Figure 2E-G). Balanced accuracy is an important metric in the context of imbalanced datasets, where one of the classes has many times more instances than the other (in the case of HEIST-DiC, No PAD cases were 14 times more frequent, and in NHANES, 10 times). Thus, the balanced accuracy represents the average accuracy obtained on both classes (PAD and No PAD), which in turn ensures that our model is equally adept at identifying both conditions.

Moreover, the ML framework also achieved high F1 scores of 0.89 for HEIST-DiC and 0.93 for NHANES, pointing to a strong balance between precision (the proportion of true positive results in all positive predictions), and recall (the proportion of true positive results in all actual positives). Thus, the F1 score in combination with the balanced accuracy provides realistic measures of a model’s strength, particularly in datasets where one class might significantly outnumber the other. In summary, the ML framework was effective in predicting the occurrence of PAD across different cohorts, even in the face of data imbalance. Further, sex and iron biomarkers were ranked among the top predictive markers for the presence of PAD in the HEIST-DiC (ferritin, iron and transferrin) and the NHANES cohorts (ferritin, TIBC, TSAT and iron; Supplementary Figure 2; Figure 3A-B).

Regression analyses

medRxiv preprint doi: https://doi.org/10.1101/2023.08.21.23293418; this version posted November 12, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

We next performed regression analyses on the HEIST-DiC data and tested the interaction effect of each of the iron biomarkers with sex and found that ferritin showed a positive association in females with PAD (Ferritin*Sex, OR 2.68, 95% CI 0.94-7.61, *p*=0.057) while, iron (Iron*Sex, OR 1.17, 95% CI 0.44-2.96, p=0.7), transferrin (Transferrin*Sex, OR 1.28, 95% CI 0.48-3.17, *p*=0.6) and TSAT (TSAT*Sex, OR 1.1, 95% CI 0.35-3.2, *p*=0.9) were non-significant. Consistent with our observations in the HEIST-DiC cohort, we found that ferritin also showed a significant interaction with sex in the NHANES cohort (Ferritin*Sex, OR 1.76, 95% CI 1.16-2.67, *p*=0.008). TIBC also had a significant negative interaction (TIBC*Sex, OR 0.61, 95% CI 0.4-0.92, *p*=0.018), while iron (Iron*Sex, OR 0.95, 95% CI 0.69-1.31, *p*=0.76) and TSAT (TSAT*Sex, OR 1.15, 95% CI 0.84-1.58, *p*=0.37) were nonsignificant.

In the multivariable regression analyses of the NHANES data, we found that, ferritin or sex alone do not show a statistically significant relationship with PAD. On the other hand, the interaction between ferritin and sex reveals a significant positive association for two spline components (Table 3). Specifically, ferritin showed positive and statistically significant effects in females with two positions (48-97 ng/mL: OR 14.59, 95% CI 1.6-135.93, p= 0.019; 98-169 ng/mL: OR 171.07, 95% CI 1.27-23404, p= 0.039); this suggests that the relationship between PAD and ferritin is nonlinear and that within high ferritin ranges females show a positive association with PAD. Other significant covariates include age, CRP, and smoking, all of which also positively impact the outcome. By contrast, no associations were significant for iron and TSAT (Supplementary Table 2).

Discussion

Our study provides a comparison between ML and regression analyses to demonstrate the association between iron biomarkers in the context of PAD. Taken together, our findings from both analyses show that serum iron biomarkers are associated with PAD, with ferritin demonstrating a consistent association specifically in females with the two methodologies applied.

The positive association for higher serum ferritin among females with PAD strengthens previous observations on the NHANES 1999-2001 cohort[33]. When this observation is compared to other related disease contexts (e.g. carotid/coronary atherosclerosis/cardiovascular disease), ferritin's positive association agrees with some studies[11,15,16], but not others[34–36] . Our findings also agree with The Iron and Atherosclerosis Study (FeAST; https://clinicaltrials.gov, NCT00032357), a major trial that investigated the benefits of iron reduction on the clinical outcomes in a cohort of individuals with PAD[37]. While the initial analysis of the FeAST trial did not report a favorable outcome following

medRxiv preprint doi: https://doi.org/10.1101/2023.08.21.23293418; this version posted November 12, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

phlebotomy[37], the reanalysis of the results adjusting for the effects of age[38] and smoking[39] have suggested that iron reduction (with ferritin and TSAT as endpoints) could be of significant benefit for younger individuals and smokers. However, the FeAST trial did not contain sufficient participants to study sex-specific differences. Our study also advances current knowledge. We extend the analyses by Menke et al[33], by including additional participants from the NHANES 2003-2004 cycle in the analysis. In contrast to the FeAST trial[37], we have included participants with prediabetes, adjusted more stringently for inflammation [excluded samples with features of inflammation (ferritin>400 ng/mL or CRP>3 mg/dL) and used CRP as a covariate] and increased statistical power to study sex-specific differences.

If the high ferritin values observed in our study were solely interpreted as an indicator of body iron stores, the data suggest that there is an increased risk of PAD at the upper end of, what is clinically accepted as “physiological range”. However, ferritin is also an acute-phase protein and inflammation influences most iron parameters. During inflammatory states, iron export via ferroportin from macrophages and duodenal enterocytes is reduced by hepcidin-controlled[40] and transcriptional mechanisms[41,42]. Consequently, macrophages become iron loaded contributing to elevated circulating ferritin levels and hypoferremia (low iron levels in the plasma)[43]. Therefore, an alternative interpretation of our findings is that elevated ferritin may be an epiphenomenon secondary to inflammation. In addition, the presence of microinflammation in tissues cannot be completely excluded from the available data.

To summarize, our study adds to the body of evidence suggesting sex-specific differences in PAD[44] and highlights a possible role of iron (directly or indirectly) in this relationship. We have partly controlled for hormonal influences by excluding NHANES participants who are currently pregnant, menstruating, lactating, or on hormonal contraceptives, but the hormone levels were not available to be included as covariates. It is further possible that additional sex-specific mechanisms are in play.

In both cohorts analyzed, we observed a similar prevalence of PAD (~6%) associated with well-known risk factors (old age, diabetes, and hypertension). Markers related to metabolic and vascular health, such as blood pressure, plasma glucose levels and CRP, were elevated among PAD participants in both cohorts. Additionally, BMI was higher in PAD participants compared to non-PAD participants in HEIST-DiC, while NHANES shows more consistent BMI values across groups, with a minor increase only in females with PAD. It is important to note that the cohorts analyzed differed in that the HEIST-DiC cohort consisted of mainly European individuals with diabetes and prediabetes from a hospital-based study, while

medRxiv preprint doi: https://doi.org/10.1101/2023.08.21.23293418; this version posted November 12, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

NHANES is a survey cohort, representative of the US population, and therefore contains a mix of ethnicities.

The strength of our study lies in the analyses of these two independent cohorts by orthogonal approaches, allowing for a generalization of the findings, an increased robustness of the data, and a reduction of cohort-specific biases. The examination of multiple iron biomarkers, including serum iron, ferritin, TSAT, and TIBC provides a comprehensive analysis of the relationship between iron and clinically apparent PAD. To the best of our knowledge, this is the first study to evaluate the relationship between iron biomarkers and PAD in prediabetes and diabetes. Our approach also underscores the complementary nature of ML and regression analyses: while ML excels at predictive accuracy and feature selection, regression analyses offer insights into interaction effects and the nuanced relationship between predictors and outcomes. Therefore, integrating both methodologies can leverage the strengths of each approach to provide a comprehensive understanding of complex disease dynamics. Limitations include the cross-sectional nature of the study which limits causal inferences and precludes the assessment of temporal relationships. Further, we were unable to control for all possible confounding factors, such as genetic predisposition, lifestyle, medications, or other comorbidities. The study described here also did not investigate potential mechanisms underlying the observed associations.

Conclusion

Overall, our data show that elevated ferritin within the physiological range is associated with clinically apparent PAD in females but not males. Further studies are necessary to confirm whether the observed associations are indeed causal and to explore the potential mechanisms underlying this relationship. Our findings need to be considered explorative to encourage the investigation of potential biological pathways underlying the sex-specific differences in iron's influence on atherosclerosis.

Ethics statement

The ethics committees of the University of Heidelberg approved this study (Decision No. 204/2004, 400/2010, and S-383/2016) according to the Declaration of Helsinki. All participants entered the study according to the guidelines of the local ethics committees following written informed consent to participate.

Funding

medRxiv preprint doi: https://doi.org/10.1101/2023.08.21.23293418; this version posted November 12, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

The HEIST-DiC study was funded by the Deutsche Forschungsgemeinschaft (SFB1118). MUM acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG) (SFB1118; GRK 2727; FerrOs - FOR5146; Ferroptosis SPP2306: Project No.461704553) and from the Federal Ministry of Education and Research (NephrESA Nr 031L0191C; Translational Lung Research Center Heidelberg (TLRC-H), German Center for Lung Research (DZL); Project No. FKZ 82DZL004A1).

Author contributions

Anand Ruban Agarvas: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing- Original draft, Visualization

Stefan Kopf: Investigation, Resources, Writing - Review & Editing

Tiago J.S Lopes: Formal analysis, Investigation, Writing- Original draft, Visualization

Paul Thalmann: Validation, Writing - Review & Editing

José Manuel Fernández-Real: Methodology, Writing - Review & Editing, Supervision

Peter Nawroth: Project administration, Funding acquisition, Supervision, Writing - Review & Editing

Martina U. Muckenthaler: Conceptualization, Methodology, Supervision, Writing - Review & Editing, Project administration, Funding acquisition

Conflict of interest

None to declare

Data and Code Availability Statement

The data from this study can be explored here: https://anandr.shinyapps.io/NHANES_HEISTDiC/. The HEIST-DiC dataset analyzed in the current study is available from the corresponding authors on reasonable request. The NHANES dataset is publicly available. Codes used for data analysis and visualization are accessible here (https://github.com/griffindoc/pad)

Funding

References

1. Aday AW, Matsushita K. Epidemiology of Peripheral Artery Disease and Polyvascular Disease. *Circ Res*. 2021;128(12):1818-1832. doi:10.1161/CIRCRESAHA.121.318535

medRxiv preprint doi: https://doi.org/10.1101/2023.08.21.23293418; this version posted November 12, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

2. Fowkes FGR, Rudan D, Rudan I, Aboyans V, Denenberg JO, McDermott MM, Norman PE, Sampson UK, Williams LJ, Mensah GA, Criqui MH. Comparison of global estimates of prevalence and risk factors for peripheral artery disease in 2000 and 2010: a systematic review and analysis. *The Lancet*. 2013;382(9901):1329-1340. doi:10.1016/S0140-6736(13)61249-0

3. Galy B, Conrad M, Muckenthaler M. Mechanisms controlling cellular and systemic iron homeostasis. *Nat Rev Mol Cell Biol*. Published online October 2, 2023. doi:10.1038/s41580-023-00648-1

4. Sullivan JeromeL. IRON AND THE SEX DIFFERENCE IN HEART DISEASE RISK. *The Lancet*. 1981;317(8233):1293-1294. doi:10.1016/S0140-6736(81)92463-6

5. Bai T, Li M, Liu Y, Qiao Z, Wang Z. Inhibition of ferroptosis alleviates atherosclerosis through attenuating lipid peroxidation and endothelial dysfunction in mouse aortic endothelial cell. *Free Radic Biol Med*. 2020;160:92-102. doi:10.1016/j.freeradbiomed.2020.07.026

6. Moreau KL, Stauffer BL, Kohrt WM, Seals DR. Essential Role of Estrogen for Improvements in Vascular Endothelial Function With Endurance Exercise in Postmenopausal Women. *The Journal of Clinical Endocrinology & Metabolism*. 2013;98(11):4507-4515. doi:10.1210/jc.2013-2183

7. Taddei S, Virdis A, Ghiadoni L, Mattei P, Sudano I, Bernini G, Pinto S, Salvetti A. Menopause Is Associated With Endothelial Dysfunction in Women. *Hypertension*. 1996;28(4):576-582. doi:10.1161/01.HYP.28.4.576

8. Yang Q, Jian J, Katz S, Abramson SB, Huang X. 17β-Estradiol Inhibits Iron Hormone Hepcidin Through an Estrogen Responsive Element Half-Site. *Endocrinology*. 2012;153(7):3170-3178. doi:10.1210/en.2011-2045

9. Li X, Rhee DK, Malhotra R, Mayeur C, Hurst LA, Ager E, Shelton G, Kramer Y, McCulloh D, Keefe D, Bloch KD, Bloch DB, Peterson RT. Progesterone receptor membrane component-1 regulates hepcidin biosynthesis. *Journal of Clinical Investigation*. 2015;126(1):389-401. doi:10.1172/JCI83831

10. Vinchi F, Porto G, Simmelbauer A, Altamura S, Passos ST, Garbowski M, Silva AMN, Spaich S, Seide SE, Sparla R, Hentze MW, Muckenthaler MU. Atherosclerosis is aggravated by iron overload and ameliorated by dietary and pharmacological iron restriction. *Eur Heart J*. Published online March 20, 2019. doi:10.1093/eurheartj/ehz112

11. Wolff B, Völzke H, Lüdemann J, Robinson D, Vogelgesang D, Staudt A, Kessler C, Dahm JB, John U, Felix SB. Association between high serum ferritin levels and carotid atherosclerosis in the study of health in Pomerania (SHIP). *Stroke*. 2004;35(2):453-457. doi:10.1161/01.STR.0000114875.31599.1C

12. Sung KC, Kang SM, Cho EJ, Park JB, Wild SH, Byrne CD. Ferritin Is Independently Associated With the Presence of Coronary Artery Calcium in 12 033 Men. *ATVB*. 2012;32(10):2525-2530. doi:10.1161/ATVBAHA.112.253088

13. Kiechl S, Willeit J, Egger G, Poewe W, Oberhollenzer F. Body iron stores and the risk of carotid atherosclerosis: prospective results from the Bruneck study. *Circulation*. 1997;96(10):3300-3307. doi:10.1161/01.cir.96.10.3300

medRxiv preprint doi: https://doi.org/10.1101/2023.08.21.23293418; this version posted November 12, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

14. Das De S, Krishna S, Jethwa A. Iron status and its association with coronary heart disease: systematic review and meta-analysis of prospective studies. *Atherosclerosis*. 2015;238(2):296-303. doi:10.1016/j.atherosclerosis.2014.12.018

15. Guo S, Mao X, Li X, Ouyang H. Association between iron status and incident coronary artery disease: a population based-cohort study. *Sci Rep*. 2022;12(1):17490. doi:10.1038/s41598-022-22275-0

16. Klip IjT, Voors AA, Swinkels DW, Bakker SJL, Kootstra-Ros JE, Lam CS, Van Der Harst P, Van Veldhuisen DJ, Van Der Meer P. Serum ferritin and risk for new-onset heart failure and cardiovascular events in the community: Serum ferritin and risk for new-onset HF and CV events. *Eur J Heart Fail*. 2017;19(3):348-356. doi:10.1002/ejhf.622

17. Gill D, Brewer CF, Monori G, Trégouët DA, Franceschini N, Giambartolomei C, INVENT Consortium, Tzoulaki I, Dehghan A. Effects of Genetically Determined Iron Status on Risk of Venous Thromboembolism and Carotid Atherosclerotic Disease: A Mendelian Randomization Study. *J Am Heart Assoc*. 2019;8(15):e012994. doi:10.1161/JAHA.119.012994

18. Altamura S, Kopf S, Schmidt J, Müdder K, da Silva AR, Nawroth P, Muckenthaler MU. Uncoupled iron homeostasis in type 2 diabetes mellitus. *J Mol Med*. 2017;95(12):1387-1398. doi:10.1007/s00109-017-1596-3

19. Zoller H, Schaefer B, Vanclooster A, Griffiths B, Bardou-Jacquet E, Corradini E, Porto G, Ryan J, Cornberg M. EASL Clinical Practice Guidelines on haemochromatosis. *Journal of Hepatology*. 2022;77(2):479-502. doi:10.1016/j.jhep.2022.03.033

20. Qiu Z, Chen X, Geng T, Wan Z, Lu Q, Li L, Zhu K, Zhang X, Liu Y, Lin X, Chen L, Shan Z, Liu L, Pan A, Liu G. Associations of Serum Carotenoids With Risk of Cardiovascular Mortality Among Individuals With Type 2 Diabetes: Results From NHANES. *Diabetes Care*. 2022;45(6):1453-1461. doi:10.2337/dc21-2371

21. Newsome PN, Cramb R, Davison SM, Dillon JF, Foulerton M, Godfrey EM, Hall R, Harrower U, Hudson M, Langford A, Mackie A, Mitchell-Thain R, Sennett K, Sheron NC, Verne J, Walmsley M, Yeoman A. Guidelines on the management of abnormal liver blood tests. *Gut*. 2018;67(1):6-19. doi:10.1136/gutjnl-2017-314924

22. Ahlqvist E, Prasad RB, Groop L. Subtypes of Type 2 Diabetes Determined From Clinical Parameters. *Diabetes*. 2020;69(10):2086-2093. doi:10.2337/dbi20-0001

23. Aboyans V, Ricco JB, Bartelink MLEL, Björck M, Brodmann M, Cohnert T, Collet JP, Czerny M, De Carlo M, Debus S, Espinola-Klein C, Kahan T, Kownator S, Mazzolai L, Naylor AR, Roffi M, Röther J, Sprynger M, Tendera M, Tepe G, Venermo M, Vlachopoulos C, Desormais I, ESC Scientific Document Group, Widimsky P, Kolh P, Agewall S, Bueno H, Coca A, De Borst GJ, Delgado V, Dick F, Erol C, Ferrini M, Kakkos S, Katus HA, Knuuti J, Lindholt J, Mattle H, Pieniazek P, Piepoli MF, Scheinert D, Sievert H, Simpson I, Sulzenko J, Tamargo J, Tokgozoglu L, Torbicki A, Tsakountakis N, Tuñón J, De Ceniga MV, Windecker S, Zamorano JL, Windecker S, Aboyans V, Agewall S, Barbato E, Bueno H, Coca A, Collet JP, Coman IM, Dean V, Delgado V, Fitzsimons D, Gaemperli O, Hindricks G, Iung B, Juni P, Katus HA, Knuuti J, Lancellotti P, Leclercq C, McDonagh T, Piepoli MF, Ponikowski P, Richter DJ, Roffi M, Shlyakhto E, Simpson IA, Zamorano JL, Zelveian PH, Haumer M, Isachkin D, De Backer T, Dilic M, Petrov I, Kirhmajer MV, Karetova D, Prescott E, Soliman H, Paapstel A, Makinen K, Tosev S, Messas E, Pagava Z, Müller OJ, Naka KK, Járai Z, et al. 2017 ESC Guidelines on the Diagnosis and Treatment of Peripheral Arterial Diseases, in

medRxiv preprint doi: https://doi.org/10.1101/2023.08.21.23293418; this version posted November 12, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

collaboration with the European Society for Vascular Surgery (ESVS). *European Heart Journal*. 2018;39(9):763-816. doi:10.1093/eurheartj/ehx095

24. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. *jair*. 2002;16:321-357. doi:10.1613/jair.953

25. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM; 2016:785-794. doi:10.1145/2939672.2939785

26. Lundberg S, Lee SI. A Unified Approach to Interpreting Model Predictions. Published online 2017. doi:10.48550/ARXIV.1705.07874

27. R Core Team. R: A Language and Environment for Statistical Computing. Published online 2022. https://www.R-project.org/

28. Rossum G van, Drake FL. *The Python Language Reference*. Release 3.0.1 [Repr.]. Python Software Foundation; 2010.

29. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011;12(85):2825-2830.

30. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*. 2007;370(9596):1453-1457. doi:10.1016/S0140-6736(07)61602-X

31. Allaire J. quarto: R Interface to "Quarto" Markdown Publishing System_. R package version 1.2. Published online 2022. https://CRAN.R-project.org/package=quarto

32. Chang W, Cheng J, Allaire J, Sievert C, Schloerke B, Xie Y, Allen J, McPherson J, Dipert A, Borges B. Shiny: Web Application Framework for R_. R package version 1.7.4.1. Published online 2023. https://CRAN.R-project.org/package=shiny

33. Menke A, Fernández-Real JM, Muntner P, Guallar E. The association of biomarkers of iron status with peripheral arterial disease in US adults. *BMC Cardiovasc Disord*. 2009;9(1):34. doi:10.1186/1471-2261-9-34

34. Zhou J, Zhao R, Wang D, Gao Q, Zhao D, Ouyang B, Hao L, Peng X. Sex-Specific Association Between Iron Status and the Predicted 10-Year Risk for Atherosclerotic Cardiovascular Disease in Hypertensive Patients. *Biol Trace Elem Res*. 2022;200(11):4594-4607. doi:10.1007/s12011-021-03060-y

35. Ellervik C, Marott JL, Tybjærg-Hansen A, Schnohr P, Nordestgaard BG. Total and cause-specific mortality by moderately and markedly increased ferritin concentrations: general population study and metaanalysis. *Clin Chem*. 2014;60(11):1419-1428. doi:10.1373/clinchem.2014.229013

36. Kadoglou NPE, Biddulph JP, Rafnsson SB, Trivella M, Nihoyannopoulos P, Demakakos P. The association of ferritin with cardiovascular and all-cause mortality in community-dwellers: The

medRxiv preprint doi: https://doi.org/10.1101/2023.08.21.23293418; this version posted November 12, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

English longitudinal study of ageing. *PLoS One*. 2017;12(6):e0178994. doi:10.1371/journal.pone.0178994

37. Zacharski LR, Chow BK, Howes PS, Shamayeva G, Baron JA, Dalman RL, Malenka DJ, Ozaki CK, Lavori PW. Reduction of Iron Stores and Cardiovascular Outcomes in Patients With Peripheral Arterial Disease: A Randomized Controlled Trial. *JAMA*. 2007;297(6):603. doi:10.1001/jama.297.6.603

38. Zacharski LR, Shamayeva G, Chow BK. Effect of controlled reduction of body iron stores on clinical outcomes in peripheral arterial disease. *American Heart Journal*. 2011;162(5):949-957.e1. doi:10.1016/j.ahj.2011.08.013

39. DePalma RG, Zacharski LR, Chow BK, Shamayeva G, Hayes VW. Reduction of iron stores and clinical outcomes in peripheral arterial disease: outcome comparisons in smokers and non-smokers. *Vascular*. 2013;21(4):233-241. doi:10.1177/1708538113478776

40. Nemeth E, Tuttle MS, Powelson J, Vaughn MB, Donovan A, Ward DM, Ganz T, Kaplan J. Hepcidin regulates cellular iron efflux by binding to ferroportin and inducing its internalization. *Science*. 2004;306(5704):2090-2093. doi:10.1126/science.1104742

41. Guida C, Altamura S, Klein FA, Galy B, Boutros M, Ulmer AJ, Hentze MW, Muckenthaler MU. A novel inflammatory pathway mediating rapid hepcidin-independent hypoferremia. *Blood*. 2015;125(14):2265-2275. doi:10.1182/blood-2014-08-595256

42. Marques O, Weiss G, Muckenthaler MU. The role of iron in chronic inflammatory diseases: from mechanisms to treatment options in anemia of inflammation. *Blood*. 2022;140(19):2011-2023. doi:10.1182/blood.2021013472

43. Cohen LA, Gutierrez L, Weiss A, Leichtmann-Bardoogo Y, Zhang D liang, Crooks DR, Sougrat R, Morgenstern A, Galy B, Hentze MW, Lazaro FJ, Rouault TA, Meyron-Holtz EG. Serum ferritin is derived primarily from macrophages through a nonclassical secretory pathway. *Blood*. 2010;116(9):1574-1584. doi:10.1182/blood-2009-11-253815

44. Pabon M, Cheng S, Altin SE, Sethi SS, Nelson MD, Moreau KL, Hamburg N, Hess CN. Sex Differences in Peripheral Artery Disease. *Circ Res*. 2022;130(4):496-511. doi:10.1161/CIRCRESAHA.121.320702

medRxiv preprint doi: https://doi.org/10.1101/2023.08.21.23293418; this version posted November 12, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Legends

Table 1. Demographics of HEIST-DiC participants

The demographics of the HEIST-DiC participants classified by sex and the presence of PAD are shown in Table 1. For continuous variables, the summary data are shown as Mean±Standard Deviation and for categorical variables as N (%). P-values as calculated by the Wilcoxon rank sum test or Pearson's Chi-squared test[C] and significant P-values are boldfaced.

BMI: Body Mass Index, HDL: High-Density Lipoprotein, LDL: Low-Density Lipoprotein, HbA1c: Glycated hemoglobin, ICAM1: Intercellular Adhesion Molecule 1, VCAM1: Vascular Cell Adhesion Molecule 1, hsCRP: High sensitivity C-reactive protein, IL6: Interleukin 6, hsTNT: High sensitivity Troponin T, proBNP: N-Terminal B-type Natriuretic Peptide, eGFR: Estimated Glomerular Filtration Rate, TSAT: Transferrin Saturation

Table 2. Demographics of NHANES 1999-2004 participants

The demographics of the NHANES 1999-2004 participants classified by sex and the presence of PAD are shown in Table 2. The [1]N of participants shown are unweighted numbers. For continuous variables, the summary data are shown as Median (Interquartile Range), and for categorical variables as N (%)- survey weights were included in calculating the summary data. P-values as calculated by the Wilcoxon rank-sum test for complex survey samples or chi-squared test with Rao & Scott's second-order correction[C] and significant P-values are boldfaced.

BMI: Body Mass Index, HDL: High-Density Lipoprotein, LDL: Low-Density Lipoprotein, HbA1c: Glycated hemoglobin, CRP: High sensitivity C-reactive protein, eGFR: Estimated Glomerular Filtration Rate, TSAT: Transferrin Saturation, TIBC: Total Iron Binding Capacity, AST: Aspartate Aminotransferase, ALT: Alanine Aminotransferase, FIB4 Score: Fibrosis-4 index Score

Table 3. Multivariable regression analysis of ferritin and PAD in the NHANES Cohort

We conducted multivariable regression analyses by flexibly modelling the probability of PAD's presence with ferritin by spline regression (degree of freedom=4). The following variables were included a covariates: age, CRP, HDLc, triacylglycerols, HbA1c, ACR, BMI, ethnicity (non-Hispanic whites as the reference), smoking (non-smokers as reference), the presence of diabetes, AST, ALT, Hb, creatinine, antihyperlipidemic medications; the former seven variables as continuous and the rest as categorical). All continuous variables were mean-centered for the analysis. In addition, sample weights were included in the analyses

BMI: Body Mass Index, HDL: High-Density Lipoprotein, CRP: High sensitivity C-reactive protein, AST: Aspartate Aminotransferase, ALT: Alanine Aminotransferase; ACR: Albumin-creatinine ratio, OR: Odds Ratio, 95% CI: 95% Confidence Interval, SE: Standard Error

Figure 1. Outline of the study and analysis approach

Figure 2. Setup and results of the machine learning approach

(A) The datasets were divided using either cross-fold validation (NHANES), or the leave-one-out strategy (HEIST-DIC). Next, the categorical variables were transformed into binary values (the so-called 'hot-encode'), we filled missing values and oversampled the minority classes (see Methods). After training the classifier, the instances of the test set had their classes predicted, and we evaluated the results using classical ML metrics. (B-D) The ROC and precision-recall curves of the HEIST-DIC cohort data, and its confusion matrix. (E-G) The ROC and precision-recall curves of the NHANES cohort data, and its confusion matrix.

Figure 3. Feature importance of machine learning classification

(A-B) We used the SHAP method to quantify the importance of the features that were contributing the most to the prediction of PAD occurrence. Here, higher values indicate a

medRxiv preprint doi: https://doi.org/10.1101/2023.08.21.23293418; this version posted November 12, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

higher contribution of a given feature. In both panels, iron-related markers are color-coded, and appear among the top features in terms of relevance to the prediction of PAD.

Supplementary files

Supplementary Table 1. Data fields and variables from NHANES 1999-2004.
The table shows the variables and the corresponding codes used for accessing data from the NHANES. The hyperlinks to the procedure manuals and datafiles are also provided in the table.

Supplementary Table 2. Multivariable regression analysis of iron biomarkers and PAD in the NHANES Cohort
We conducted multivariable regression analyses by flexibly modelling the probability of PAD's presence with iron biomarkers (iron or TSAT) by spline regression (degree of freedom=4). The following variables were included aas covariates: age, CRP, HDLc, triacylglycerols, HbA1c, ACR, BMI, ethnicity (non-Hispanic whites as the reference), smoking (non-smokers as reference), the presence of diabetes, AST, ALT, Hb, creatinine, antihyperlipidemic medications; the former seven variables as continuous and the rest as categorical). All continuous variables were mean-centered for the analysis. In addition, sample weights were included in the analyses
BMI: Body Mass Index, HDL: High-Density Lipoprotein, CRP: High sensitivity C-reactive protein, AST: Aspartate Aminotransferase, ALT: Alanine Aminotransferase; ACR: Albumin-creatinine ratio, OR: Odds Ratio, 95% CI: 95% Confidence Interval, SE: Standard Error

Supplementary Figure 1. Feature importance of xGBoost classification
(A-B) We used the XGBoost classifier and its output to quantify the importance of the features to the prediction of PAD occurrence. In both panels, higher values indicate a higher contribution of a given feature, and iron-related markers are color coded, and appear among the top features in terms of relevance to the prediction of this condition.

	Males				Females			
Variable	**Overall** N = 192[1]	**No PAD** N = 175[1]	**PAD** N = 17[1]	**p-value**[2]	**Overall** N = 165[1]	**No PAD** N = 159[1]	**PAD** N = 6[1]	**p-value**[2]
Age (years)	63 (54, 69)	62 (54, 69)	70 (66, 76)	**<0.001**	60 (53, 67)	59 (52, 67)	69 (62, 73)	**0.025**
BMI (kg/sq.m)	29.0 (25.9, 32.9)	29.0 (25.9, 33.0)	29.7 (26.2, 32.9)	0.6	29.4 (24.9, 33.0)	29.3 (24.8, 33.0)	32.8 (30.5, 40.4)	**0.036**
Hypertension								
No	70 (36%)	67 (38%)	3 (18%)	0.2	82 (50%)	82 (52%)	0 (0%)	**0.039**
Yes	122 (64%)	108 (62%)	14 (82%)		83 (50%)	77 (48%)	6 (100%)	
Diabetes								
No diabetes	37 (19%)	37 (21%)	0 (0%)	0.11	67 (41%)	66 (42%)	1 (17%)	0.2
Prediabetes	19 (9.9%)	17 (9.7%)	2 (12%)		18 (11%)	18 (11%)	0 (0%)	
Diabetes	136 (71%)	121 (69%)	15 (88%)		80 (48%)	75 (47%)	5 (83%)	
Dietary management								
No	170 (89%)	156 (89%)	14 (82%)	0.7	155 (94%)	150 (94%)	5 (83%)	0.8
Yes	22 (11%)	19 (11%)	3 (18%)		10 (6.1%)	9 (5.7%)	1 (17%)	
Oral antidiabetic medications use								
No	125 (65%)	115 (66%)	10 (59%)	0.8	122 (74%)	118 (74%)	4 (67%)	>0.9
Yes	67 (35%)	60 (34%)	7 (41%)		43 (26%)	41 (26%)	2 (33%)	
Statins use								

medRxiv preprint doi: https://doi.org/10.1101/2023.08.21.23293418; this version posted November 12, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

No	129 (67%)	118 (67%)	11 (65%)	>0.9	136 (82%)	132 (83%)	4 (67%)	0.6
Yes	63 (33%)	57 (33%)	6 (35%)		29 (18%)	27 (17%)	2 (33%)	
Insulin use								
No	157 (82%)	145 (83%)	12 (71%)	0.4	141 (85%)	137 (86%)	4 (67%)	0.5
Yes	35 (18%)	30 (17%)	5 (29%)		24 (15%)	22 (14%)	2 (33%)	
Fasting blood glucose (mg/dL)	127 (103, 161)	125 (102, 154)	167 (123, 204)	**0.002**	106 (92, 145)	105 (92, 139)	152 (139, 174)	0.067
HbA1c (g/dL)	6.40 (5.70, 7.50)	6.40 (5.60, 7.40)	7.50 (6.30, 7.90)	**0.014**	5.80 (5.40, 6.80)	5.80 (5.40, 6.80)	7.10 (6.60, 8.20)	0.056
Parameters of vascular function								
Ankle-brachial index (Right)	1.10 (1.03, 1.16)	1.11 (1.06, 1.16)	0.86 (0.76, 0.88)	**<0.001**	1.08 (1.02, 1.15)	1.09 (1.03, 1.15)	0.82 (0.76, 0.86)	**<0.001**
Ankle-brachial index (Left)	1.10 (1.05, 1.16)	1.11 (1.06, 1.17)	0.92 (0.88, 0.95)	**<0.001**	1.09 (1.04, 1.15)	1.09 (1.05, 1.15)	0.97 (0.86, 0.98)	**<0.001**
Carotid intima-media thickness (mm)	0.60 (0.54, 0.70)	0.60 (0.54, 0.70)	0.70 (0.59, 0.85)	**0.025**	0.60 (0.50, 0.66)	0.58 (0.50, 0.65)	0.83 (0.70, 0.98)	**0.002**
Carotid-femoral Pulse Wave Velocity (m/s)	8.70 (7.70, 9.80)	8.70 (7.60, 9.70)	9.75 (9.55, 11.50)	**0.004**	8.10 (7.10, 9.70)	8.10 (7.10, 9.60)	11.35 (11.00, 12.60)	**0.002**
ICAM1 (ng/mL)	294 (212, 366)	289 (206, 357)	380 (303, 419)	**0.008**	305 (225, 363)	305 (229, 367)	225 (218, 363)	0.6
VCAM1 (ng/mL)	847 (692, 1,052)	827 (682, 1,033)	1,104 (878,	**0.017**	804 (697, 961)	803 (697, 962)	856 (467, 905)	0.8

medRxiv preprint doi: https://doi.org/10.1101/2023.08.21.23293418; this version posted November 12, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

			1,189)					
Other laboratory parameters								
Triacylglycerols (mg/dL)	128 (90, 199)	125 (87, 191)	177 (148, 247)	**0.011**	119 (82, 166)	118 (80, 163)	177 (128, 207)	0.092
Cholesterol (mg/dL)	184 (157, 221)	184 (156, 221)	183 (159, 216)	0.8	204 (173, 231)	204 (173, 231)	184 (159, 240)	0.5
HDL (mg/dL)	46 (39, 55)	47 (39, 55)	40 (38, 47)	0.067	57 (47, 71)	58 (47, 71)	50 (37, 57)	0.3
LDL (mg/dL)	107 (82, 134)	107 (82, 135)	107 (83, 132)	0.9	116 (89, 139)	116 (90, 139)	81 (78, 116)	0.088
Hemoglobin (g/dL)	14.90 (14.10, 15.50)	14.90 (14.10, 15.50)	14.30 (13.20, 15.10)	0.14	13.60 (13.00, 14.20)	13.60 (13.00, 14.20)	13.25 (12.80, 14.00)	0.5
Mean Corpuscular Volume (fL)	89.0 (86.0, 92.0)	89.0 (86.0, 92.0)	90.0 (87.0, 92.0)	0.2	88.0 (85.0, 91.0)	88.0 (85.0, 91.0)	87.5 (84.0, 91.0)	0.7
hsCRP (mg/L)	1.09 (0.52, 2.54)	1.05 (0.51, 2.35)	1.32 (0.79, 4.36)	0.3	2.1 (0.7, 4.7)	1.9 (0.7, 4.6)	5.0 (2.9, 15.1)	0.083
IL6 (pg/mL)	2.45 (2.00, 3.70)	2.35 (2.00, 3.60)	3.20 (2.40, 4.20)	**0.010**	2.55 (2.00, 4.15)	2.50 (2.00, 4.10)	3.80 (2.50, 6.90)	0.3
hsTNT (ng/L)	9.0 (7.0, 13.0)	9.0 (7.0, 12.0)	11.0 (9.0, 20.0)	**0.033**	6.0 (4.0, 8.0)	6.0 (4.0, 8.0)	9.0 (8.0, 9.0)	0.2
proBNP (pg/mL)	60 (31, 129)	59 (29, 118)	113 (35, 251)	**0.052**	71 (39, 130)	69 (37, 121)	175 (135, 236)	**0.008**
eGFR (ml/min/1.73m^2)	91 (80, 100)	92 (82, 101)	76 (62, 91)	**0.002**	93 (81, 101)	93 (82, 101)	64 (50, 86)	**0.016**

medRxiv preprint doi: https://doi.org/10.1101/2023.08.21.23293418; this version posted November 12, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Urinary Albumin Creatinine Ratio (mg/g)	8 (4, 23)	8 (4, 18)	25 (11, 139)	**0.001**	7 (4, 14)	7 (4, 13)	26 (14, 111)	**0.012**
Iron-related parameters								
Iron (μM)	95 (75, 119)	96 (75, 120)	85 (76, 108)	0.7	90 (71, 118)	90 (70, 118)	95 (79, 126)	0.6
Ferritin (ng/mL)	155 (81, 259)	153 (78, 250)	192 (126, 306)	0.10	79 (44, 136)	77 (42, 131)	141 (90, 278)	**0.031**
Transferrin (g/L)	2.38 (2.18, 2.61)	2.38 (2.18, 2.61)	2.24 (2.15, 2.73)	0.6	2.42 (2.24, 2.69)	2.42 (2.24, 2.67)	2.56 (2.16, 2.88)	0.6
TSAT (%)	25 (20, 32)	26 (20, 33)	23 (21, 27)	0.2	24 (18, 31)	25 (18, 31)	22 (19, 24)	0.5

[1]Median (Q1, Q3); n (%)

Table 1. Demographics of HEIST-DiC participants

The demographics of the HEIST-DiC participants classified by sex and the presence of PAD are shown in Table 1. For continuous variables, the summary data are shown as Mediann (Q1, Q3) and for categorical variables as N (%). P-values as calculated by the Wilcoxon rank sum test or Pearson's Chi-squared test[C] and significant P-values are boldfaced.

BMI: Body Mass Index, HDL: High-Density Lipoprotein, LDL: Low-Density Lipoprotein, HbA1c: Glycated hemoglobin, ICAM1: Intercellular Adhesion Molecule 1, VCAM1: Vascular Cell Adhesion Molecule 1, hsCRP: High sensitivity C-reactive protein, IL6: Interleukin 6, hsTNT: High sensitivity Troponin T, proBNP: N-Terminal B-type Natriuretic Peptide, eGFR: Estimated Glomerular Filtration Rate, TSAT: Transferrin Saturation

medRxiv preprint doi: https://doi.org/10.1101/2023.08.21.23293418; this version posted November 12, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

	Males				Females			
Characteristic	**Overall N = 2792[1]**	**No PAD, N = 2573[1]**	**1, N = 219[1]**	**p-value[2]**	**Overall N =2309[1]**	**0, N = 2082[1]**	**1, N = 227[1]**	**p-value[2]**
Ethnicity								
Non-Hispanic White	1459 (78%)	1330 (78%)	129 (83%)	**0.003**	1089 (73%)	971 (72%)	118 (76%)	**<0.001**
Non-Hispanic Black	497 (9.0%)	454 (8.9%)	43 (11%)		475 (11%)	418 (11%)	57 (16%)	
Mexican-American	662 (5.6%)	619 (5.6%)	43 (5.2%)		546 (5.0%)	504 (5.1%)	42 (4.2%)	
Others	174 (7.6%)	170 (7.9%)	4 (1.2%)		199 (11%)	189 (12%)	10 (4.0%)	
Age (years)	52 (45, 61)	51 (45, 60)	68 (60, 76)	**<0.001**	52 (45, 65)	51 (45, 64)	70 (54, 81)	**<0.001**
BMI (kg/sq.m)	27.9 (25.1, 31.0)	27.9 (25.1, 30.9)	27.8 (25.0, 32.2)	>0.9	27 (24, 32)	27 (24, 32)	29 (25, 34)	**0.034**
Diabetes (coded)								
No diabetes	1796 (75%)	1698 (76%)	98 (49%)	**<0.001**	1461 (73%)	1358 (75%)	103 (51%)	**<0.001**
Prediabetes	444 (13%)	387 (12%)	57 (27%)		418 (16%)	357 (15%)	61 (25%)	
Diabetes	450 (12%)	395 (12%)	55 (24%)		349 (11%)	296 (10%)	53 (23%)	
Hypertension (coded)								
No	252 (14%)	233 (14%)	19 (13%)	0.9	209 (15%)	188 (15%)	21 (14%)	0.9
Yes	1557	1404	153 (87%)		1296	1129	167 (86%)	

medRxiv preprint doi: https://doi.org/10.1101/2023.08.21.23293418; this version posted November 12, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

	(86%)	(86%)			(85%)	(85%)		
Smoking (coded)								
Nonsmoker	974 (36%)	931 (37%)	43 (18%)	**<0.001**	1397 (58%)	1277 (58%)	120 (51%)	**0.089**
Former smoker	1134 (39%)	1023 (38%)	111 (49%)		532 (24%)	466 (23%)	66 (28%)	
Current smoker	680 (25%)	615 (25%)	65 (33%)		375 (18%)	335 (18%)	40 (21%)	
Alcohol drinker (coded)								
No	188 (33%)	174 (33%)	14 (39%)	0.5	552 (44%)	484 (43%)	68 (50%)	0.3
Yes	339 (67%)	317 (67%)	22 (61%)		570 (56%)	506 (57%)	64 (50%)	
Antihyperlipidemic medications (coded)								
No	2269 (91%)	2083 (91%)	186 (91%)	0.9	1984 (95%)	1788 (95%)	196 (91%)	**0.037**
Yes	177 (8.7%)	159 (8.7%)	18 (9.2%)		113 (5.2%)	97 (4.8%)	16 (9.3%)	
Antiplatelet medications (coded)								
No	2404 (99%)	2209 (99%)	195 (97%)	**0.006**	2075 (99%)	1867 (99%)	208 (99%)	0.3
Yes	42 (1.1%)	33 (1.0%)	9 (2.9%)		22 (0.8%)	18 (0.7%)	4 (1.5%)	
Ankle-brachial Index (Left)	1.16 (1.09, 1.22)	1.16 (1.10, 1.22)	0.88 (0.79, 1.01)	**<0.001**	1.11 (1.04, 1.18)	1.12 (1.06, 1.19)	0.88 (0.78, 0.94)	**<0.001**
Ankle-brachial	1.16 (1.10,	1.16 (1.11,	0.87 (0.75,	**<0.001**	1.10 (1.03,	1.11 (1.05,	0.86 (0.77,	**<0.001**

medRxiv preprint doi: https://doi.org/10.1101/2023.08.21.23293418; this version posted November 12, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Index (Right)	1.22)	1.22)	0.97)		1.18)	1.18)	0.90)	
Albumin (g/L)	44.00 (42.00, 46.00)	44.00 (42.00, 46.00)	43.00 (41.00, 44.00)	**<0.001**	43.0 (41.0, 44.0)	43.0 (41.0, 45.0)	42.0 (39.0, 44.0)	**0.001**
AST (U/L)	26 (21, 33)	26 (21, 34)	22 (18, 27)	**<0.001**	19 (15, 25)	19 (15, 25)	18 (16, 23)	0.2
ALT (U/L)	24 (21, 29)	24 (21, 29)	23 (20, 28)	0.059	21 (18, 25)	21 (18, 25)	22 (19, 26)	0.15
Blood urea nitrogen (mmol/L)	5.00 (4.28, 6.10)	5.00 (3.93, 6.07)	6.07 (4.64, 7.14)	**<0.001**	4.60 (3.57, 5.71)	4.60 (3.57, 5.71)	5.00 (3.90, 6.78)	**0.002**
Creatinine (μmol/L)	88 (71, 97)	80 (71, 97)	88 (80, 106)	**0.018**	62 (53, 71)	62 (53, 71)	71 (62, 88)	**<0.001**
Total Cholesterol (mg/dL)	203 (179, 228)	204 (179, 228)	193 (171, 220)	**0.008**	207 (184, 235)	207 (183, 235)	211 (188, 238)	0.3
Triacylglycerols (mg/dL)	132 (89, 200)	132 (89, 200)	130 (86, 190)	>0.9	108 (73, 160)	107 (71, 156)	132 (95, 195)	**<0.001**
LDL (mg/dL)	124 (103, 147)	124 (103, 148)	115 (85, 136)	**0.038**	122 (101, 145)	122 (101, 145)	119 (97, 143)	0.8
HDL (mg/dL)	44 (37, 54)	44 (37, 54)	44 (37, 52)	0.4	55 (46, 67)	56 (46, 67)	52 (44, 61)	**0.006**
Fasting plasma glucose (mg/dL)	101 (94, 111)	100 (94, 110)	108 (92, 125)	0.3	95 (90, 104)	95 (90, 103)	102 (96, 115)	**<0.001**
Insulin (pmol/L)	57 (37, 92)	56 (37, 90)	85 (41, 129)	0.089	48 (33, 80)	47 (32, 77)	99 (46, 130)	**0.007**
CRP (mg/dL)	0.18 (0.08, 0.37)	0.18 (0.08, 0.37)	0.27 (0.14, 0.44)	**<0.001**	0.25 (0.10, 0.58)	0.25 (0.10, 0.57)	0.32 (0.15, 0.76)	**0.036**

medRxiv preprint doi: https://doi.org/10.1101/2023.08.21.23293418; this version posted November 12, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Hemoglobin (g/dL)	15.30 (14.70, 16.00)	15.30 (14.70, 16.00)	15.00 (14.10, 15.80)	**0.006**	13.90 (13.10, 14.50)	13.90 (13.10, 14.50)	13.80 (12.90, 14.70)	>0.9
FIB4 Score	1.14 (0.89, 1.48)	1.13 (0.89, 1.47)	1.28 (1.02, 1.62)	**0.002**	0.85 (0.63, 1.15)	0.84 (0.62, 1.14)	1.07 (0.80, 1.37)	**<0.001**
Albumin-creatinine ratio (μg/mg)	0.59 (0.40, 1.06)	0.58 (0.40, 1.01)	1.07 (0.63, 3.09)	**<0.001**	0.81 (0.54, 1.54)	0.79 (0.54, 1.44)	1.30 (0.70, 2.67)	**<0.001**
Iron (μM)	16.5 (12.9, 20.8)	16.7 (12.9, 20.9)	15.4 (11.6, 20.4)	0.065	14.1 (10.6, 18.1)	14.3 (10.6, 18.1)	12.9 (10.2, 16.5)	**0.025**
Ferritin (ng/mL)	142 (93, 223)	142 (93, 223)	120 (81, 197)	0.3	60 (31, 109)	60 (31, 107)	85 (56, 192)	**<0.001**
TSAT (%)	26 (20, 33)	27 (20, 33)	25 (17, 32)	0.079	22 (16, 29)	22 (16, 29)	20 (16, 28)	0.2
TIBC (μM)	63 (57, 70)	63 (57, 69)	63 (57, 70)	0.8	64 (58, 71)	64 (58, 71)	61 (54, 70)	**0.001**

Table 2. Demographics of NHANES 1999-2004 participants

The demographics of the NHANES 1999-2004 participants classified by sex and the presence of PAD are shown in Table 2. The [1]N of participants shown are unweighted numbers. For continuous variables, the summary data are shown as Median (Q1, Q3), and for categorical variables as N (%)- survey weights were included in calculating the summary data. P-values as calculated by the Wilcoxon rank-sum test for complex survey samples or chi-squared test with Rao & Scott's second-order correction[C] and significant P-values are boldfaced.

BMI: Body Mass Index, HDL: High-Density Lipoprotein, LDL: Low-Density Lipoprotein, HbA1c: Glycated hemoglobin, CRP: High sensitivity C-reactive protein, eGFR: Estimated Glomerular Filtration Rate, TSAT: Transferrin Saturation, TIBC: Total Iron Binding Capacity, AST: Aspartate Aminotransferase, ALT: Alanine Aminotransferase, FIB4 Score: Fibrosis-4 index Score

medRxiv preprint doi: https://doi.org/10.1101/2023.08.21.23293418; this version posted November 12, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

medRxiv preprint doi: https://doi.org/10.1101/2023.08.21.23293418; this version posted November 12, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Table 3	**OR**	**SE**	**Z score**	**P. Value**	**95% CI**	
					Low	**High**
Intercept	0.001	3.205	-2.338	**0.019**	0.000	0.297
Ferritin (<48 ng/mL)	0.868	0.896	-0.158	0.875	0.150	5.031
Ferritin (48-96 ng/mL)	0.481	1.478	-0.495	0.621	0.027	8.714
Ferritin (97-168 ng/mL)	0.284	2.034	-0.619	0.536	0.005	15.306
Ferritin (>169 ng/mL)	0.777	1.061	-0.238	0.812	0.097	6.214
Sex:F (ref=M)	0.281	1.053	-1.205	0.228	0.036	2.215
Ferritin (<48 ng/mL)*Females	1.042	1.177	0.035	0.972	0.104	10.471
Ferritin (48-96 ng/mL) *Females	14.593	1.139	2.354	**0.019**	1.567	135.935
Ferritin (97-168 ng/mL) *Females	171.066	2.501	2.056	**0.040**	1.270	23034.375
Ferritin (>169 ng/mL) *Females	3.073	1.247	0.900	0.368	0.266	35.427
Former Smoker (ref= Never Smokers)	1.319	0.392	0.706	0.480	0.612	2.844
Current Smoker (ref= Never Smokers)	4.696	0.282	5.487	**<0.001**	2.703	8.160
Prediabetes (ref=No diabetes)	1.570	0.257	1.754	0.080	0.948	2.600
Diabetes Mellitus (ref=No diabetes)	1.337	0.403	0.720	0.472	0.607	2.945
Non-Hispanic Black (ref: Non-Hispanic White)	1.296	0.398	0.651	0.515	0.594	2.829
Mexican-American (ref: Non-Hispanic White)	1.281	0.311	0.797	0.425	0.697	2.355
Other Ethnicities (ref: Non-Hispanic White)	0.251	0.740	-1.865	0.062	0.059	1.072
CRP	1.810	0.259	2.293	0.022	1.090	3.004
Age	1.095	0.013	7.241	**<0.001**	1.068	1.122
Creatinine	1.009	0.006	1.448	0.148	0.997	1.021
ACR	0.058	1.987	-1.430	0.153	0.001	2.868
Triacylglycerols	0.997	0.068	-0.039	0.969	0.873	1.140
BMI	1.078	0.162	0.463	0.644	0.785	1.480
HDL	0.751	0.345	-0.831	0.406	0.382	1.476
Hemoglobin	0.897	0.190	-0.574	0.566	0.618	1.301
Antihyperlipidemic medications (ref=No)	1.318	0.546	0.505	0.613	0.452	3.841
AST	1.005	0.013	0.438	0.662	0.981	1.030
ALT	0.988	0.018	-0.647	0.518	0.953	1.024

HEIST-DiC Cohort

NHANES 1999-2004

Iron biomarkers

- *Iron*
- *Ferritin*
- *TSAT*
- *Transferrin*

Vascular function markers

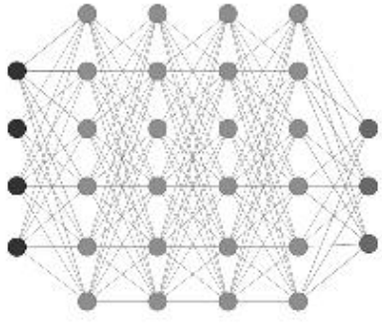
- *ICAM1*
- *VCAM1*
- *Ankle-Brachial Index*

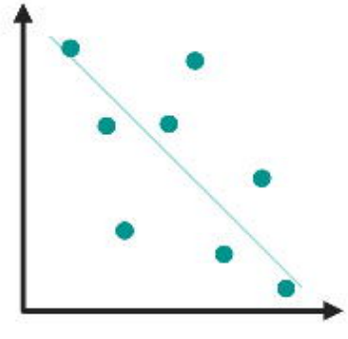
Iron biomarkers

- *Iron*
- *Ferritin*
- *TSAT*
- *TIBC*

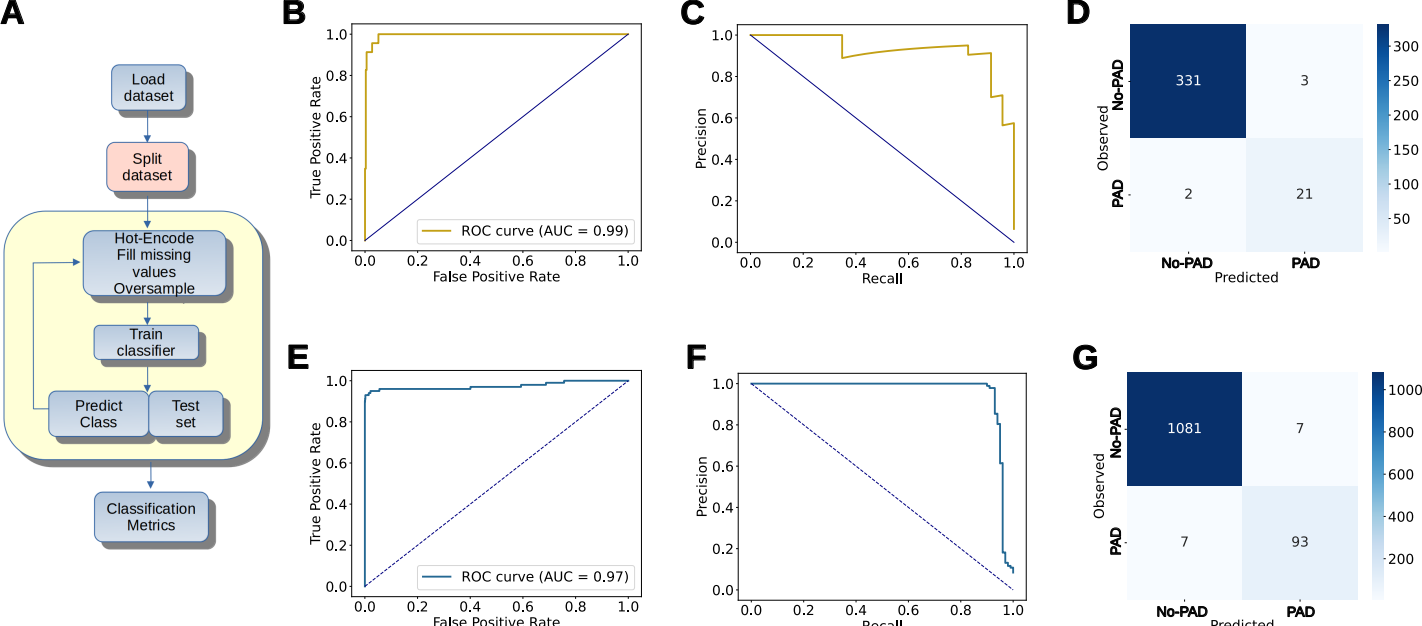
Vascular function markers

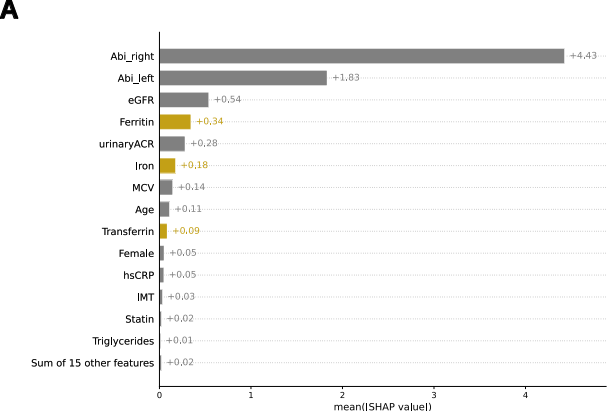
- *Ankle-Brachial Index*

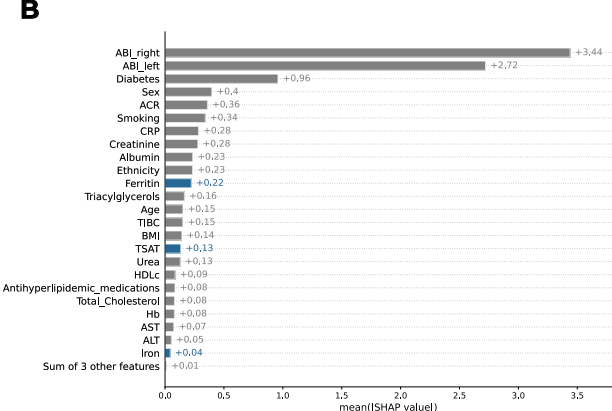




Machine learning & Regression analysis


A
Load dataset
Split dataset
Hot-Encode
Fill missing values
Oversample
Train classifier
Predict Class
Test set
Classification Metrics
B
True Positive Rate
False Positive Rate
ROC curve (AUC = 0.99)
C
Precision
Recall
D
Observed
Predicted
No-PAD
PAD
331
3
2
21
E
True Positive Rate
False Positive Rate
ROC curve (AUC = 0.97)
F
Precision
Recall
G
Observed
Predicted
No-PAD
PAD
1081
7
7
93


A
Abi_right
+4.43
Abi_left
+1.83
eGFR
+0.54
Ferritin
+0.34
urinaryACR
+0.28
Iron
+0.18
MCV
+0.14
Age
+0.11
Transferrin
+0.09
Female
+0.05
hsCRP
+0.05
IMT
+0.03
Statin
+0.02
Triglycerides
+0.01
Sum of 15 other features
+0.02
0
1
2
3
4
mean(|SHAP value|)


B
ABI_right
+3.44
ABI_left
+2.72
Diabetes
+0.96
Sex
+0.4
ACR
+0.36
Smoking
+0.34
CRP
+0.28
Creatinine
+0.28
Albumin
+0.23
Ethnicity
+0.23
Ferritin
+0.22
Triacylglycerols
+0.16
Age
+0.15
TIBC
+0.15
BMI
+0.14
TSAT
+0.13
Urea
+0.13
HDLc
+0.09
Antihyperlipidemic_medications
+0.08
Total_Cholesterol
+0.08
Hb
+0.08
AST
+0.07
ALT
+0.05
Iron
+0.04
Sum of 3 other features
+0.01
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
mean(|SHAP value|)