|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Training Set** | **Validation Set** | **Test Set** | **p-value** |
| Positive gentic test (n, %) | 147, 74.4% | 35, 71.42% | 40, 85.1% | 0.23 |
| Age (years) | 39 (32 - 50) | 40 (27 – 47) | 41 (32 – 54) | 0.21 |
| Sex (men, %) | 106, 53.53% | 30, 61.22% | 20, 42.56% | 0.18 |
| BMI (kg/m2) | 25.91 (23.26 – 28) | 25.41 (22.68 – 28.41) | 25.37 (22.08 – 27.34) | 0.38 |
| Total Cholesterol (mg/dL) | 306.7 (278 – 359.6) | 321 (281.9 – 372.2) | 350.5 (297.8 – 375.1) | 0.14 |
| LDL-C (CM) (mg/dL) | 238 (203.5 – 285.9) | 251.7 (214.5 – 310.1) | 272.9 8216.2 – 315.3) | 0.10 |
| HDL-C (CM) (mg/dL) | 50 (42.15 – 60.32) | 49.71 (39.64 – 57.81) | 52.10 (45.35 – 66.80) | 0.28 |
| Triglycerides (mg/dL) | 97.43 (70.86 – 141.72) | 88.57 (62 – 129.43) | 101.86 (74.10 – 148.36) | 0.36 |
| **Lipoprotein particle number** | | | | |
| VLDL (nM) | 31.30 (21.23 – 56.48) | 40.85 (22.80 – 58.13) | 29.43 (18.30 – 37.80) | 0.16 |
| Large VLDL (nM) | 0.65 (0.41 – 1.13) | 0.85 (0.51 – 1.22) | 0.60 (0.30 – 0.80) | 0.10 |
| Medium VLDL (nM) | 3.55 (2.34 – 5.86) | 4.27 (2.66 – 6.71) | 3.23 (1.86 – 5.17) | 0.13 |
| Small VLDL (nM) | 27.54 (18.49 – 49.54) | 35.55 (19.58 – 51.04) | 25.99 (15.96 – 32.34) | 0.16 |
| LDL (nM) | 2066 (1819 – 2370) | 1984 (1746 – 2251) | 1990 (1791 – 2336) | 0.30 |
| Large LDL (nM) | 315.7 (274.3 – 358.1) | 296 (257.9 – 337.3) | 311.1 (271.9 – 369.2) | 0.16 |
| Medium LDL (nM) | 810.5 (669.0 – 993.6) | 738.4 (612.9 – 927.9) | 794.3 (662.8 – 973.0) | 0.10 |
| Small LDL (nM) | 926.6 (836.8 – 1085.1) | 927.5 (799.6 – 1054.2) | 888.8 (792.6 – 1077.4) | 0.43 |
| HDL (uM) | 26.80 (23.35 – 30.32) | 25.46 (23.53 – 29.54) | 27.57 (23.80 – 31.78) | 0.30 |
| Large HDL (uM) | 0.37 (0.33 – 0.42) | 0.36 (0.2 – 0.40) | 0.36 (0.34 – 0.44) | 0.46 |
| Medium HDL (uM) | 10.20 (9.07 – 11.21) | 10.01 (8.96 – 10.90) | 10.75 (9.38 – 11.84) | 0.17 |
| Small HDL (uM) | 16.32 (13.16 – 19.27) | 15.70 (13.70 – 18.22) | 15.87 (13.22 – 20.56) | 0.59 |
| **Lipoprotein composition (mg/dl)** | | | | |
| VLDL-C | 9.87 (4.56 – 20.58) | 10.01 (5.467 – 20.59) | 7.31 (2.60 – 11.86) | 0.11 |
| VLDL-TG | 45.90 (30.37 – 77.30) | 41.05 (28.72 – 69.28) | 37.33 (24.21 – 50.05) | 0.13 |
| IDL-C | 13.62 (10.12 – 18.31) | 14.16 (10.59 – 19.92) | 11.94 (8.74 – 18.07) | 0.43 |
| IDL-TG | 11.96 (8.70 – 15.04) | 11.80 (9.71 – 15.08) | 10.46 (6.89 – 15.14) | 0.55 |
| Remnant-C | 23.85 (16.71 – 37.10) | 22.12 (14.60 – 33.56) | 24.73 (18.21 – 38.39) | 0.52 |
| LDL-C | 206.5 (183.8 – 244) | 222.6 (197.5 – 244.6) | 215.4 (189.6 – 237.3) | 0.48 |
| LDL-TG | 25.74 (21.30 – 31.56) | 27.29 (21.96 – 31.59) | 25.13 (21.43 – 30.35) | 0.75 |
| HDL-C | 57.13 (51.57 – 63.82) | 58.08 (49.01 – 65.43) | 60.11 (51.33 – 69.04) | 0.56 |
| HDL-TG | 9.16 (5.35 – 13.62) | 9.58 (6.22 – 12.30) | 7.75 (3.32 – 14.41) | 0.61 |
| **Lipoprotein size (diameter, nm)** | | | | |
| VLDL | 42.08 (41.97 – 42.22) | 42.09 (42.01 – 42.19) | 42.09 (41.99 – 42.21) | 0.76 |
| LDL | 21.43 (21.21 – 21.57) | 21.35 (21.17 – 21.48) | 21.52 (21.35 – 21.65) | 0.01 |
| HDL | 8.33 (8.27 – 8.39) | 8.32 (8.28 – 8.37) | 8.35 (8.26 – 8.39) | 0.90 |

***Supplementary Table 1*:** Description of the training data, validation data and test data. Normally distributed data is expressed in terms of mean and standard deviation (SD). Nonnormally distributed data is expressed in terms of median and interquartile range (IQR). Categorical data is expressed in percentages. To evaluate differences between groups, one way anova was used to compare normally distributed data, Kruskal-Wallis test was used to compare nonnormally distributed data and chi square (*X*2) test was used to compare categorical data. P-values < 0.05 are considered statistically significant. BMI = body mass index, VLDL = very low density lipoprotein, LDL = low density lipoprotein, HDL = high density lipoprotein.

|  |  |  |
| --- | --- | --- |
|  | **Hyperparameters selected** | **Variables selected in order importance** |
| **Random Forest** | Mtry = 8  Number of trees = 10000 | Age  VLDL particles  Large VLDL particles  Medium VLDL particles  Small VLDL particles  VLDL-TG  HDL particle number  Large HDL particles  Small HDL particles  HDL diameter  HDL-TG  IDL-TG  LDL particle number  Large LDL particles  Medium LDL particles  LDL diameter  LDL-C  LDL-TG  Remnant-C |
| **Logistic Regression** | Lambda = 0.57 | Large LDL particles  HDL diameter  Chol-Rem |

**Supplementary Table 2:** Hyperparameters tuned and variables selected for each training model.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **INTERNAL VALIDATION** | | | | | | | | | | |
| **MODEL** | **CONFUSION MATRIX** | | | **AUC**  **(95 CI%)** | **Sensitivity**  **(95% CI)** | **Specificity**  **(95% CI)** | **PPV**  **(95% CI)** | **NPV**  **(95% CI)** | **Accuracy (95% CI)** | **F1 score** |
| **RF** |  | **-** | **+** | 0.86  **(0.66 – 0.93)** | 0.82  (0.48 – 0.98) | 0.64  (0.49 – 0.79) | 0.41  (0.21 – 0.64) | 0.92  (0.74 – 0.99) | 0.68  (0.53 – 0.81) | 0.55 |
| **-** | 9 | 13 |  |
| **+** | 2 | 23 |  |
| **LR** |  | **-** | **+** | 0.83  **(0.71 – 0.93)** | 0.91  (0.59 – 1) | 0.69  (0.52 – 0.84) | 0.47  (0.26 – 0.70) | 0.96  (0.80 – 1) | 0.74  (0.60 – 0.86) | 0.29 |
| **-** | 10 | 11 |  |
| **+** | 1 | 25 |  |
| **EXTERNAL VALIDATION** | | | | | | | | | | |
| **MODEL** | **CONFUSION MATRIX** | | | **AUC**  **(95% MD CI)** | **Sensitivity**  **(95% CI)** | **Specificity**  **(95% CI)** | **PPV**  **(95% CI)** | **NPV**  **(95% CI)** | **Accuracy (95% CI)** | **F1 score** |
| **RF** |  | **-** | **+** | 0.60  **(0.56 – 0.72)** | 0.43  (0.23 -0.66) | 0.59  (0.46 – 0.67) | 0.21  (0.10 – 0.34) | 0.81  (0.68 – 0.89) | 0.56  (0.47 – 0.65) | 0.26 |
| **-** | 10 | 37 |  |
| **+** | 13 | 54 |  |
| **LR** |  | **-** | **+** | 0.59  **(0.57 – 0.61)** | 0.56  (0.34 – 0.77) | 0.59  (0.49 – 0.70) | 0.26  (0.15 – 0.40) | 0.84  (0.73 – 0.92) | 0.59  (0.49 – 0.68) | 0.36 |
| **-** | 13 | 37 |  |
| **+** | 10 | 54 |  |

**Supplementary Table 3:** Confusion matrix and performance metrics of the adjusted models with LDL cholesterol, age, sex and BMI measured by conventional methods. - = negative genetic test, + = positive genetic test. MD = model distribution, CI = confidence interval.

C:\Users\Didac\OneDrive - URV\URLA\FH DAIANA - JOSEFA\supfig1.tif

**Supplementary Figure 1:** ROC Curves for the predictions of the internal dataset (A) and external dataset (B) with LDL conventionally measured.