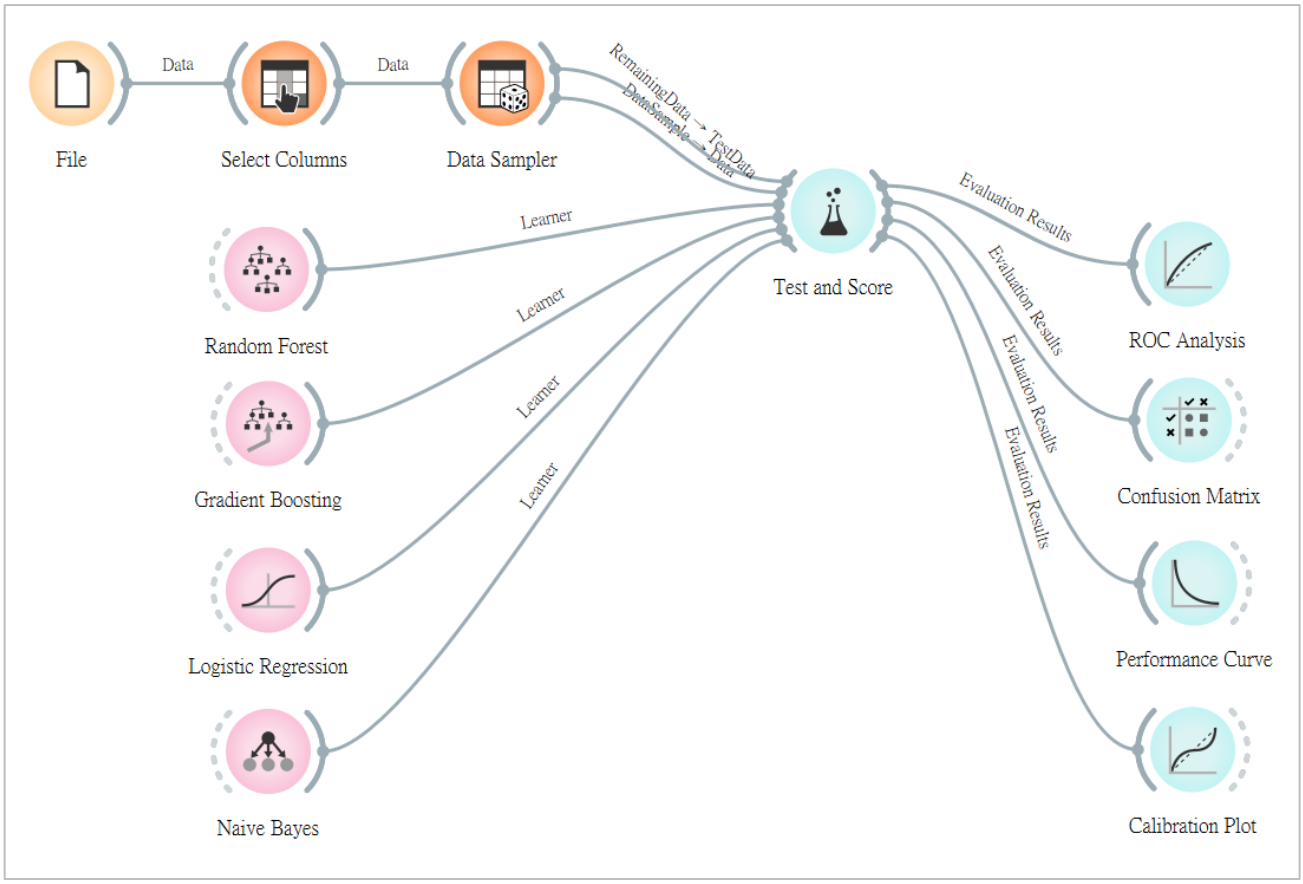

Supplementary Figure 1 Orange data mining widget settings.

Supplementary Table 1 The association between mtDNA polymorphisms and severe degree metabolic dysfunction-associated steatotic liver disease in the screening dataset

mtSNP	Screening study population					
	MASLD ≥ 3 CMRF	Control	*P* value (Fisher's or χ^2)	OR (95% CI)	Location	Call rate (%)
n	73	131				
146T>C	12(16.4%)	18(13.7%)	0.602	1.24(0.56-2.73)	D-loop	100.0
150C>T	13(17.8%)	34(26.0%)	0.185	0.62(0.30-1.26)	D-loop	100.0
152T>C	13(17.8%)	24(18.3%)	0.927	0.97(0.46-2.04)	D-loop	100.0
189A>G	0(0.0%)	1(0.8%)	1.000	0.64(0.58-0.71)	D-loop	100.0
195T>C	4(5.5%)	10(7.6%)	0.774	0.70(0.21-2.32)	D-loop	100.0
199T>C	5(6.8%)	16(12.2%)	0.227	0.53(0.19-1.51)	D-loop	100.0
249delA	9(12.5%)	35(26.9%)	0.017	0.39(0.17-0.86)	D-loop	99.0
489T>C	29(39.7%)	58(44.6%)	0.499	0.82(0.46-1.47)	D-loop	99.5
663A>G	7(9.6%)	10(7.6%)	0.628	1.28(0.47-3.53)	12S rRNA	100.0
709G>A	19(26.0%)	16(12.2%)	0.012	2.53(1.21-5.30)	12S rRNA	100.0
827A>G	3(4.1%)	6(4.6%)	1.000	0.89(0.22-3.65)	12S rRNA	99.5
1382A>C	4(5.5%)	1(0.8%)	0.058	7.42(0.81-67.69)	12S rRNA	99.0
1438A>G	71(97.3%)	130(99.2%)	0.292	0.27(0.02-3.06)	12S rRNA	100.0
1719G>A	1(1.4%)	4(3.1%)	0.656	0.44(0.05-3.99)	16S rRNA	99.5

1736A>G	7(9.6%)	10(7.8%)	0.651	1.26(0.46-3.47)	16S rRNA	99.0
3010G>A	14(19.2%)	12(9.2%)	0.042	2.33(1.02-5.36)	16S rRNA	99.5
3394T>C	0(0.0%)	1(0.8%)	1.000	0.64(0.58-0.71)	ND1(Tyr→His)	99.5
3970C>T	6(13.6%)	16(20.8%)	0.327	0.60(0.22-1.67)	ND1(Leu→Leu)	59.3
4071C>T	4(5.5%)	16(12.3%)	0.144	0.41(0.13-1.29)	ND1(Tyr→Tyr)	99.5
4048G>A	2(2.7%)	11(8.5%)	0.141	0.31(0.07-1.41)	ND1(Asp→Asn)	99.5
4216T>C	0(0.0%)	0(0.0%)	-	-	ND1(Tyr→His)	99.5
4291T>C	0(0.0%)	0(0.0%)	-	-	tRNA Ile	100.0
4386T>C	0(0.0%)	0(0.0%)	-	-	tRNA Gln	99.5
4824A>G	7(9.6%)	12(9.3%)	0.947	1.03(0.39-2.76)	ND2(Thr→Ala)	99.0
4833A>G	4(5.5%)	2(1.5%)	0.191	3.71(0.66-20.77)	ND2(Thr→Ala)	99.5
4883C>T	15(20.5%)	23(17.7%)	0.617	1.20(0.58-2.48)	ND2(Pro→Pro)	99.5
4958A>G	0(0.0%)	0(0.0%)	-	-	ND2(Met→Met)	98.5
5178C>A	15(20.5%)	23(17.7%)	0.617	1.20(0.58-2.48)	ND2(Leu→Met)	99.5
5231G>A	4(5.5%)	7(5.4%)	1.000	1.01(0.29-3.57)	ND2(Leu→Leu)	99.0
5263C>T	3(4.1%)	2(1.5%)	0.351	2.76(0.45-16.94)	ND2(Ala→Val)	100.0
5301A>G	1(1.4%)	9(7.0%)	0.097	0.18(0.02-1.48)	ND2(Ile→Val)	98.5
5460G>A	2(2.7%)	10(7.8%)	0.218	0.34(0.07-1.57)	ND2(Ala→Thr)	99.0
6392T>C	10(13.9%)	27(20.9%)	0.217	0.61(0.28-1.34)	COI(Asn→Asn)	98.5
6455C>T	4(5.5%)	11(8.4%)	0.580	0.63(0.19-2.06)	COI(Phe→Phe)	100.0

7598G>A	1(1.4%)	1(0.8%)	1.000	1.76(0.11-28.63)	COII(Ala→Thr)	98.5
8414C>T	6(20.7%)	5(9.3%)	0.143	2.56(0.71-9.25)	TAPase8(Leu→Phe)	40.7
8473C>T	1(3.4%)	1(1.9%)	1.000	1.89(0.11-31.42)	ATPase8(Thr→Thr)	40.7
8563A>G	4(5.5%)	7(5.3%)	1.000	1.03(0.29-3.63)	ATPase6(Thr→Ala)	100.0
8584G>A	11(15.1%)	21(16.0%)	0.856	0.93(0.42-2.05)	ATPase6(Ala→Thr)	100.0
8701A>G	28(38.4%)	61(46.6%)	0.257	0.71(0.40-1.28)	ATPase6(Thr→Ala)	100.0
8794C>T	7(9.6%)	9(6.9%)	0.489	1.44(0.51-4.04)	ATPase6(His→Tyr)	100.0
9053G>A	0(0.0%)	0(0.0%)	-	-	ATPase6(Ser→Asn)	100.0
9090T>C	0(0.0%)	0(0.0%)	-	-	ATPase6(Ser→Ser)	99.5
10084T>C	0(0.0%)	2(1.5%)	0.538	0.64(0.58-0.71)	ND3(Ser→Ser)	100.0
10238T>C	1(1.4%)	2(1.5%)	1.000	0.90(0.08-10.50)	ND3(Ile→Ile)	100.0
10310G>A	9(12.3%)	26(20.0%)	0.165	0.56(0.25-1.28)	ND3(Leu→Leu)	99.5
10397A>G	1(1.4%)	8(6.1%)	0.162	0.21(0.03-1.74)	ND3(Trp→Trp)	100.0
10398A>G	43(58.9%)	68(51.9%)	0.336	1.33(0.75-2.37)	ND3(Thr→Ala)	100.0
10400C>T	30(41.1%)	61(46.6%)	0.451	0.80(0.45-1.43)	ND3(Thr→Ala)	100.0
10609T>C	5(6.8%)	12(9.2%)	0.567	0.73(0.25-2.16)	ND4L(Met→Thr)	100.0
10873T>C	30(41.1%)	59(45.7%)	0.523	0.83(0.46-1.48)	ND4(Pro→Pro)	99.0
11084A>G	0(0.0%)	0(0.0%)	-	-	ND4(Thr→Ala)	99.5
12308 A>G	0(0.0%)	0(0.0%)	-	-	tRNA-Leu	99.0
12358A>G	4(5.5%)	8(6.2%)	1.000	0.88(0.26-3.04)	ND5(Thr→Ala)	99.5

12361A>G	11(15.1%)	2(1.5%)	$3.0x10^{-4}$	11.36(2.44-52.80)	ND5(Thr→Ala)	99.5
12372G>A	0(0.0%)	4(3.1%)	0.299	0.63(0.57-0.70)	ND5(Leu→Leu)	99.0
12406G>A	5(6.9%)	14(10.9%)	0.364	0.61(0.21-1.78)	ND5(Val→Ile)	98.5
12705C>T	40(54.8%)	76(58.5%)	0.612	0.86(0.48-1.54)	ND5(Ile→Ile)	99.5
12811T>C	2(2.7%)	8(6.3%)	0.333	0.42(0.09-2.05)	ND5(Tyr→His)	98.5
13105A>G	1(1.4%)	2(1.6%)	1.000	0.88(0.08-9.82)	ND5(Ile→Val)	98.5
13263A>G	0(0.0%)	5(3.8%)	0.162	0.63(0.57-0.70)	ND5(Gln→Gln)	99.5
13708G>A	3(4.1%)	11(8.7%)	0.263	0.45(0.12-1.66)	ND5(Ala→Thr)	97.5
13759G>A	3(4.1%)	8(6.2%)	0.749	0.65(0.17-2.54)	ND5(Ala→Thr)	99.5
13879T>C	0(0.0%)	0(0.0%)	-	-	ND5(Ser→Pro)	100.0
13928G>C	9(12.3%)	30(24.0%)	0.046	0.45(0.20-1.00)	ND5(Ser→Thr)	97.1
14470T>C	1(1.4%)	8(6.2%)	0.161	0.21(0.03-1.73)	ND6(Gly→Gly)	99.5
14502T>C	2(2.8%)	1(0.8%)	0.287	3.71(0.33-41.69)	ND6(Ile→Val)	99.5
14569G>A	4(5.6%)	3(2.3%)	0.250	2.49(0.54-11.45)	ND6(Ser→Ser)	99.0
14979T>C	2(2.7%)	2(1.5%)	0.620	1.80(0.25-13.07)	Cytb(Ile→Thr)	99.5
15218A>G	1(1.4%)	1(0.8%)	1.000	1.81(0.11-29.30)	Cytb(Thr→Ala)	100.0
15323G>A	1(1.4%)	1(0.8%)	1.000	1.81(0.11-29.30)	Cytb(Ala→Thr)	100.0
15497G>A	2(2.7%)	1(0.8%)	0.292	3.66(0.33-41.09)	Cytb(Gly→Ser)	100.0
16111C>T	4(5.5%)	1(0.8%)	0.059	7.36(0.81-67.17)	D-loop	98.5
16126T>C	5(6.8%)	6(4.6%)	0.500	1.52(0.45-5.16)	D-loop	99.5

16129G>A	14(19.7%)	20(15.6%)	0.462	1.33(0.62-2.82)	D-loop	97.5
16189T>C	19(27.5%)	25(19.4%)	0.188	1.58(0.80-3.14)	D-loop	97.1
16278C>T	2(2.8%)	2(1.5%)	0.617	1.83(0.25-13.26)	D-loop	99.0
16304T>C	7(9.6%)	26(20.2%)	0.051	0.42(0.17-1.02)	D-loop	99.0
16324T>C	1(1.4%)	2(1.6%)	1.000	0.88(0.08-9.90)	D-loop	99.0
16362T>C	26(35.6%)	36(27.9%)	0.254	1.43(0.77-2.64)	D-loop	99.0
16519T>C	41(56.9%)	67(52.8%)	0.569	1.18(0.66-2.12)	D-loop	97.5
8272del9bp	23(31.5%)	20(15.3%)	0.006	2.55(1.29-5.07)	MinNCR10	100.0

Severe degree MASLD, metabolic dysfunction-associated steatotic liver disease with more than 3 cardiometabolic risk factors; mtDNA, mitochondrial DNA

Supplementary Table 2 The association between common mtDNA haplogroups and severe degree metabolic dysfunction-associated steatotic liver disease in the screening dataset

Haplogroup	**MASLD ≥ 3 CMRF**	**Control**	***P* value (Fisher's or χ^2)**	**Odds ratio (95% CI)**
n	73	131		
A	7(9.6%)	9(7.1%)	0.530	1.39(0.50-3.91)
B	23(31.5%)	19(15.0%)	0.006	2.62(1.31-5.23)
C	0(0.0%)	4(3.1%)	0.299	0.63(0.56-0.70)
D4	8(11.0%)	8(6.3%)	0.242	1.83(0.66-5.11)
D5	1(1.4%)	8(6.3%)	0.159	0.21(0.03-1.69)

E	1(1.4%)	1(0.8%)	1.000	1.75(0.11-28.40)
F	9(12.3%)	26(20.5%)	0.145	0.55(0.24-1.24)
G	2(2.7%)	0(0.0%)	0.132	0.36(0.30-0.43)
M7a	0(0.0%)	0(0.0%)	-	-
M7b	3(4.1%)	12(9.4%)	0.264	0.41(0.11-1.51)
M8a	0(0.0%)	9(7.1%)	0.028	0.62(0.55-0.69)
M9	0(0.0%)	0(0.0%)	-	-
M10	3(4.1%)	1(0.8%)	0.139	5.40(0.55-52.90)
M27	0(0.0%)	2(1.6%)	0.534	0.63(0.57-0.70)
N9a	2(2.7%)	6(4.7%)	0.713	0.57(0.11-2.89)
R9	0(0.0%)	5(3.9%)	0.161	0.63(0.56-0.70)
R11	1(1.4%)	1(0.8%)	1.000	1.75(0.11-28.40)
R30	1(1.4%)	0(0.0%)	0.365	0.36(0.30-0.44)
Z	0(0.0%)	0(0.0%)	-	-

Severe degree MASLD, metabolic dysfunction-associated steatotic liver disease with more than 3 cardiometabolic risk factors; mtDNA, mitochondrial DNA.

Supplementary Table 3 Mitochondrial DNA haplogroup classification for Asian population

Mitochondrial DNA Haplogroup classification for Asian Samples

H	6	7	1	1	3	3	4	4	4	4	4	5	5	7	82	8	8	8	8	9	1	1	1	1	1	1	1	1	1	1	1	1	1
G	6	0	7	7	0	3	0	8	8	8	9	1	3	5	72	4	5	7	7	0	0	0	0	0	0	0	2	2	3	3	4	4	5	6	6
s	3	9	1	3	1	9	7	2	3	8	5	7	0	9	de	1	8	0	9	9	3	3	3	4	8	3	7	2	9	4	5	2	1	5
			9	6	0	4	1	4	3	3	8	8	1	8	l9	4	4	1	4	0	1	9	9	0	7	5	0	6	2	7	6	1	8	1
																					0	7	8	0	3	8	5	3	8	0	9	8	9	9
A	G	…	…	G	…	…	…	G	…	…	…	…	…	…	…	…	…	A	T	…	…	…	…	C	T	…	…	…	…	…	…	…	…	…
B	…	…	…	…	…	…	…	…	…	…	…	…	…	…	P	…	…	A	…	…	…	…	…	C	T	…	C	…	…	…	…	…	C	C
C	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	A	…	…	…	…	…	…	T	…	…	…	G	…	…	…	…	…	…
D	…	…	…	…	A	…	…	…	…	T	…	A	…	…	…	T	…	…	…	…	…	…	…	T	…	…	…	…	…	…	…	…	…	…
4																																		
D	…	…	…	…	…	…	…	…	…	T	…	A	G	…	…	…	…	…	…	…	…	G	…	T	…	…	…	…	…	…	…	…	…	…
5																																		
E	…	…	…	…	…	…	…	…	…	…	…	…	…	A	…	…	…	…	…	…	…	…	…	T	…	…	…	…	…	…	…	…	…	…

F	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	A	…	…	A	…	…	C	T	…	C	…	C	…	…	…	…	…
G	…	A	…	…	…	…	…	…	G	…	…	…	…	…	…	…	…	…	…	…	…	…	…	T	…	…	…	…	…	…	A	…	…	…
M7a	…	…	…	…	…	…	…	…	…	…	G	…	…	…	…	…	…	…	…	…	…	…	…	T	…	…	…	…	…	…	…	…	…	…
M7b	…	…	…	…	…	…	T	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	T	…	…	…	…	…	…	…	…	…	…
M8a	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	A	…	…	…	…	…	…	T	…	…	…	…	…	C	…	…	…	…
M9	…	…	…	…	…	C	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	T	…	…	…	…	…	…	…	…	…	…
M10	…	A	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	T	…	…	…	…	…	…	…	G	…	…

M27	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	T	…	G	…	…	…	…	…	…	…	…
N9a	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	A	…	…	…	…	…	C	T	G	…	…	…	…	…	…	…	…
R11	…	A	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	A	…	…	…	…	G	C	T	…	C	…	…	…	…	…	…	…
R30	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	A	A	…	…	…	…	…	C	T	…	C	…	…	…	…	…	…	…
R9	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	A	…	…	…	…	…	C	T	…	C	…	C	…	…	…	…	…
Z	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	A	…	…	C	…	…	…	T	…	…	…	…	…	…	…	…	…	…

HGs: Haplogroups.

Supplementary Table 4 The input features in the machine learning models

	Features
Sociodemographic	Sex, age, body mass index, waist circumference, hip circumference
Comorbidity	Diabetes mellitus, hypertension, dyslipidemia, cardiometabolic risk factors
Biochemistry	Alanine aminotransferase ULN ratio
	Aspartate aminotransferase ULN ratio
	Fasting blood sugar
	Cholesterol
	Triglyceride
	High-density Lipoprotein Cholesterol
	Low-density Lipoprotein Cholesterol
Gene	mt12361A>G

ULN: Upper limit of normal.

Supplementary Table 5 Hyperparameter tuning in the artificial intelligence models

Algorithms	Hyperparameter	Hyperparameter tuning	Final model
Random forest	number of trees	100, 200, 300, 400, 500	300
	maximal depth	3, 4, 5, 6, 7, 8	5
	minimal samples split	1, 2, 3, 4, 5	2
	minimal samples leaf	1, 2, 3	1

XGBoost	number of trees	100, 150, 200, 250, 300	150
	learning rate	0.1, 0.2, 0.3, 0.4, 0.5	0.1
	Regularization (lambda)	0.1, 1, 10, 100	lambda=1
	maximal depth	3, 4, 5, 6, 7	3
	fraction of training instances	0.5, 0.6, 0.7, 0.8, 0.9, 1	1
	fraction of features for each tree	0.5, 0.6, 0.7, 0.8, 0.9, 1	1
	fraction of features for each level	0.5, 0.6, 0.7, 0.8, 0.9, 1	1
	fraction of feature for each split	0.5, 0.6, 0.7, 0.8, 0.9, 1	1
Logistic regression	regularization type	Lasso (L1), Ridge (L2)	Ridge (L2)
	Strength (C)	0.1, 1, 10, 100	C=1
Naïve Bayes	default	-	default

Supplementary Table 6 Confusion matrix of the artificial intelligence models

Algorithms		**Actual MASLD**	**Actual Control**	**Accuracy**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Training	Predicted							
Random forest	MASLD	411	0	100%	100%	100%	100%	100%
	Control	0	506					
XGBoost	MASLD	380	44	91.8%	92.5%	91.3%	89.6%	93.7%
	Control	31	462					

Naïve Bayes	MASLD	330	151	74.7%	80.3%	70.2%	68.6%	81.4%
	Control	81	355					
Logistic regression	MASLD	281	104	74.5%	68.4%	79.4%	73.0%	75.6%
	Control	130	402					
Validation								
Random forest	MASLD	136	48	80.2%	81.9%	78.9%	73.9%	85.6%
	Control	30	179					
XGBoost	MASLD	126	53	76.3%	75.9%	76.7%	70.4%	81.3%
	Control	40	174					
Naïve Bayes	MASLD	135	57	77.6%	81.3%	74.9%	70.3%	84.6%
	Control	31	170					
Logistic regression	MASLD	112	42	75.6%	67.5%	81.5%	72.7%	77.4%
	Control	54	185					

PPV: Positive predictive value; NPV: Negative predictive value.

Supplementary Table 7 Delong test

***P* value (z-score)**	**XGBoost**	**Naïve Bayes**	**Logistic regression**
Training			

Random Forest	< 0.001 (6.34)	< 0.001 (13.04)	< 0.001 (13.22)
XGBoost	-	< 0.001 (12.49)	< 0.001 (12.51)
Naïve Bayes		-	0.441 (0.77)
Logistic regression			-
Validation			
Random forest	0.180 (1.34)	0.011 (2.53)	0.096 (1.66)
XGBoost	-	0.372 (0.89)	0.610 (0.51)
Naïve Bayes		-	0.806 (-0.25)
Logistic regression			-

SUPPLEMENTARY MATERIAL

AI model construction

Platform: Orange data mining (https://orangedatamining.com/).

Orange is a data mining tool for data scientists. Users can focus on data analysis rather than tedious coding, simplifying the construction of AI models. In Orange, data analysis is done by stacking components into workflows. Each component, called a widget, embeds data retrieval, preprocessing, visualization, modeling, or evaluation tasks. Combining different widgets in a workflow enables users to build comprehensive data analysis schemas.

Feature selection

Orange provides information gain, Gini decrease, ANOVA, and X^2 methods for feature selection. After the importance ranking, 17 features were added into the AI models. (Supplementary Table 4)

Input data

A total of 1,310 participants were randomly assigned to the training (70%) and validation (30%) datasets. Seventeen features, including sociodemographic variables, comorbidities, biochemical data, and mt12361A>G variants, were input into the machine learning models (Supplemental Table 4).

Algorithms

Random forest: A random forest (RF) is an ensemble ML approach used for regression and classification tasks. The ensemble technique combines multiple weak classifiers to obtain a better predictive performance. The random forest builds multiple decision trees via bootstrapping and combines their predictions to make a final decision by majority voting (classification task) or averaging (regression task). The random forest approach involves variability among individual trees, diminishing overfitting and enhancing predictive accuracy.

XGBoost: The eXtreme gradient boosting (XGBoost) technique is an ensemble technique that uses a gradient-boosting framework. XGBoost builds upon an ensemble of decision trees and sequentially adds new trees to correct the errors made by previous models. It employs a gradient descent method to minimize loss when creating new models. Theoretically, each new model fits new observations more precisely, thus improving overall accuracy.

Logistic regression: Logistic regression is a supervised ML algorithm used for classification tasks. It uses a sigmoid function to map predicted values to probabilities. The sigmoid function produces a probability value between 0 and 1.

Naïve bayes: Naive Bayes is a classification algorithm based on Bayes' Theorem, with the assumption of independence among features. It describes the probability of an event based on prior knowledge of conditions related to the event.

P(A | B) is the posterior probability of class A given the features B.

Parameter tuning processes

In Orange, users can manually adjust the optimal values for the model's hyperparameters, and the system will automatically display the performance metrics of the model. The ranges of hyperparameters for each algorithm were listed in the supplementary Table 5.

Model performance metrics

Confusion matrix:

		Predicted		
		Positive	**Negative**	
Ground Truth	**Positive**	True Positive (TP)	False Negative (FN) [Type II Error]	**Sensitivity (Recall)** $\frac{TP}{(TP + FN)}$
	Negative	False Positive (FP) [Type I Error]	True Negative (TN)	**Specificity** $\frac{TN}{(TN + FP)}$
		Precision (PPV) $\frac{TP}{(TP + FP)}$	**NPV** $\frac{TN}{(TN + FN)}$	**Accuracy** $\frac{TP + TN}{(TP + TN + FP + FN)}$

Prediction metrics	Definition
Sensitivity	The proportion of correctly predicted positive individuals.
Specificity	The proportion of correctly predicted negative individuals.
Positive predictive value (PPV)	The proportion of true positive results among all positive results (both true positives and false positives) predicted by the model.
Negative predictive value (NPV)	The proportion of true negative results among all negative results (both true negatives and false negatives) predicted by the model.
Accuracy	The proportion of correctly classified instances out of all cases.
Precision	The proportion of true positive predictions out of all positive predictions made by the model. It indicates the model's ability to avoid false positives.
Recall	The proportion of true positive predictions from all actual positive instances in the dataset. It indicates the model's ability to capture all positive instances and avoid false negatives.
F1 score	The F1 score is the harmonic mean of precision and recall. It provides a balanced assessment of a model's performance, particularly in an imbalanced dataset.
Area under the receiver operating characteristic (AUROC)	The ROC curve visualizes the trade-off between true positive and false positive rates for different classification thresholds. The AUROC provides a single scalar value that represents the discriminative ability across all possible thresholds.