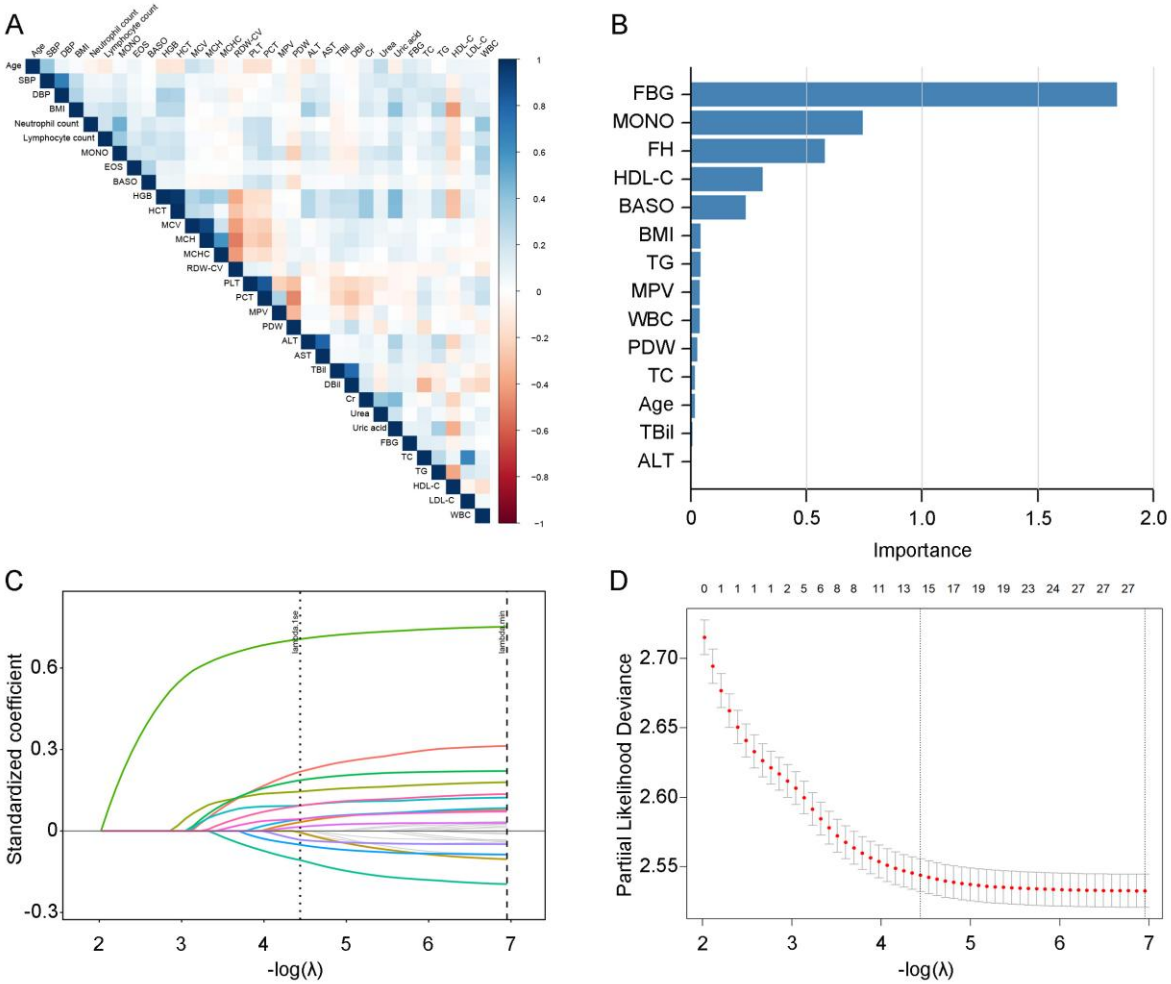
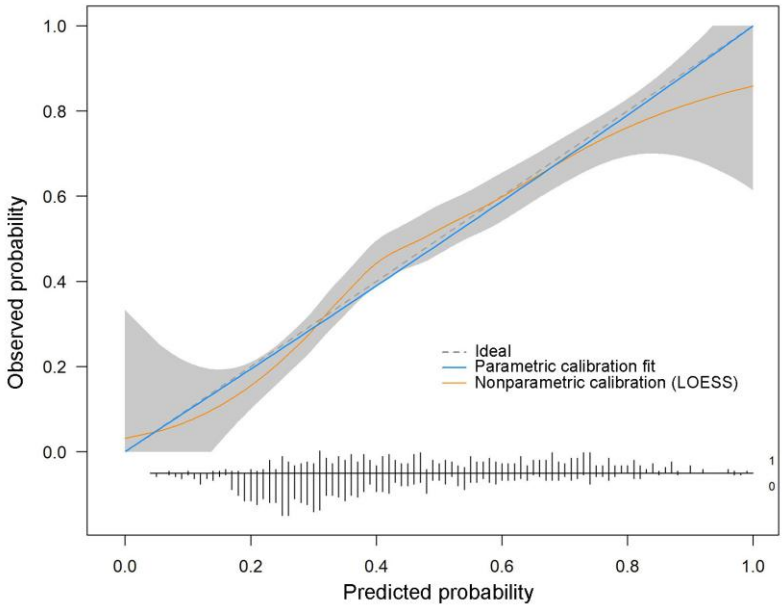
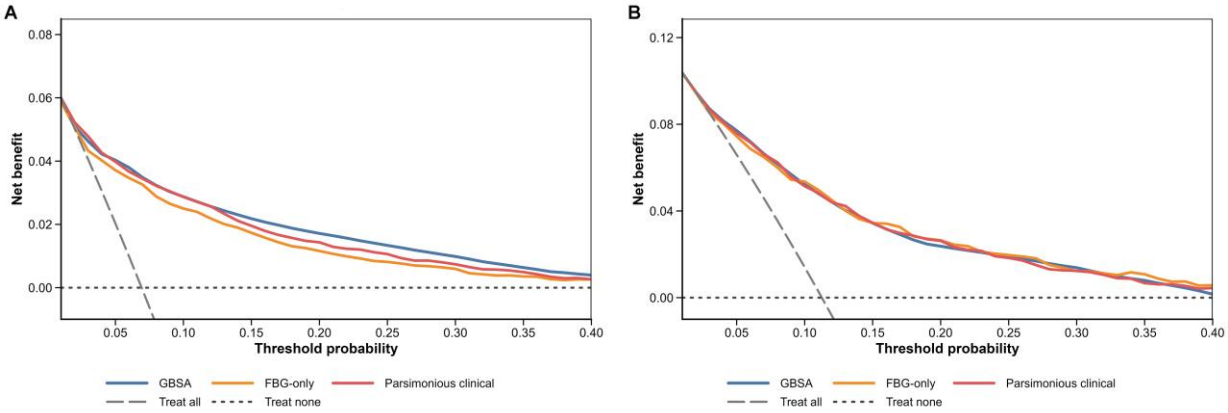


Supplementary Figure 1 Feature selection for model development. A: Heatmap of Spearman correlation coefficients among candidate variables; B: Feature selection using least absolute shrinkage and selection operator Cox regression, showing the relative importance of candidate predictors; C: Standardized coefficient path plot showing the trajectories of candidate predictors across values of -log(lambda); D: Cross-validation error plot for selecting the penalty parameter. LASSO: Least absolute shrinkage and selection operator; SBP: Systolic blood pressure; DBP: Diastolic blood pressure; BMI: Body mass index; MONO: Monocyte count; EOS: Eosinophil count; BASO: Basophil count; HGB: Hemoglobin; HCT: Hematocrit; MCV: Mean corpuscular volume; MCH: Mean corpuscular hemoglobin; MCHC: Mean corpuscular hemoglobin concentration; RDW-CV: Red blood cell distribution width-coefficient; PLT: Platelets; PCT: Plateletcrit; MPV: Mean platelet volume; PDW: Platelet distribution width; ALT: Alanine aminotransferase; AST: Aspartate aminotransferase;

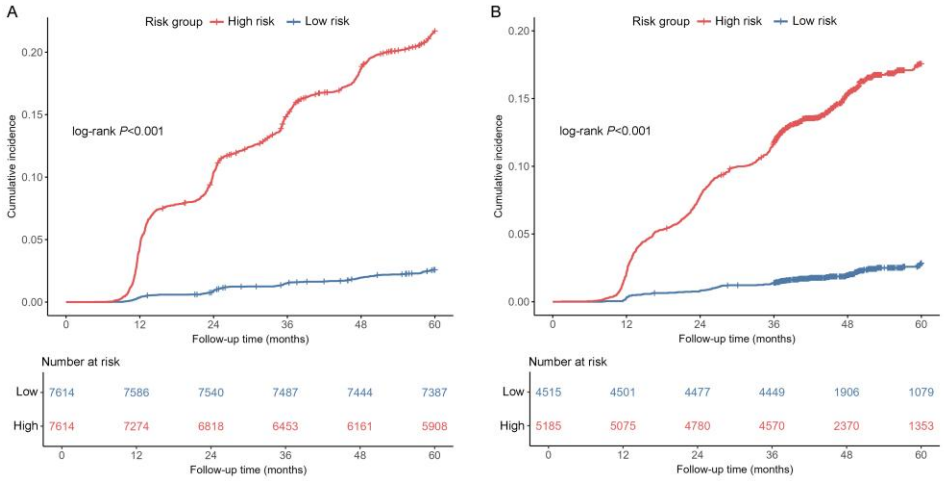
TBil: Total bilirubin; DBil: Direct bilirubin; Cr: Creatinine; FBG: Fasting blood glucose; TC: Total cholesterol; TG: Triglycerides; HDL-C: High-density lipoprotein cholesterol; LDL-C: Low-density lipoprotein cholesterol; WBC: White blood cell count; FH: Family history.



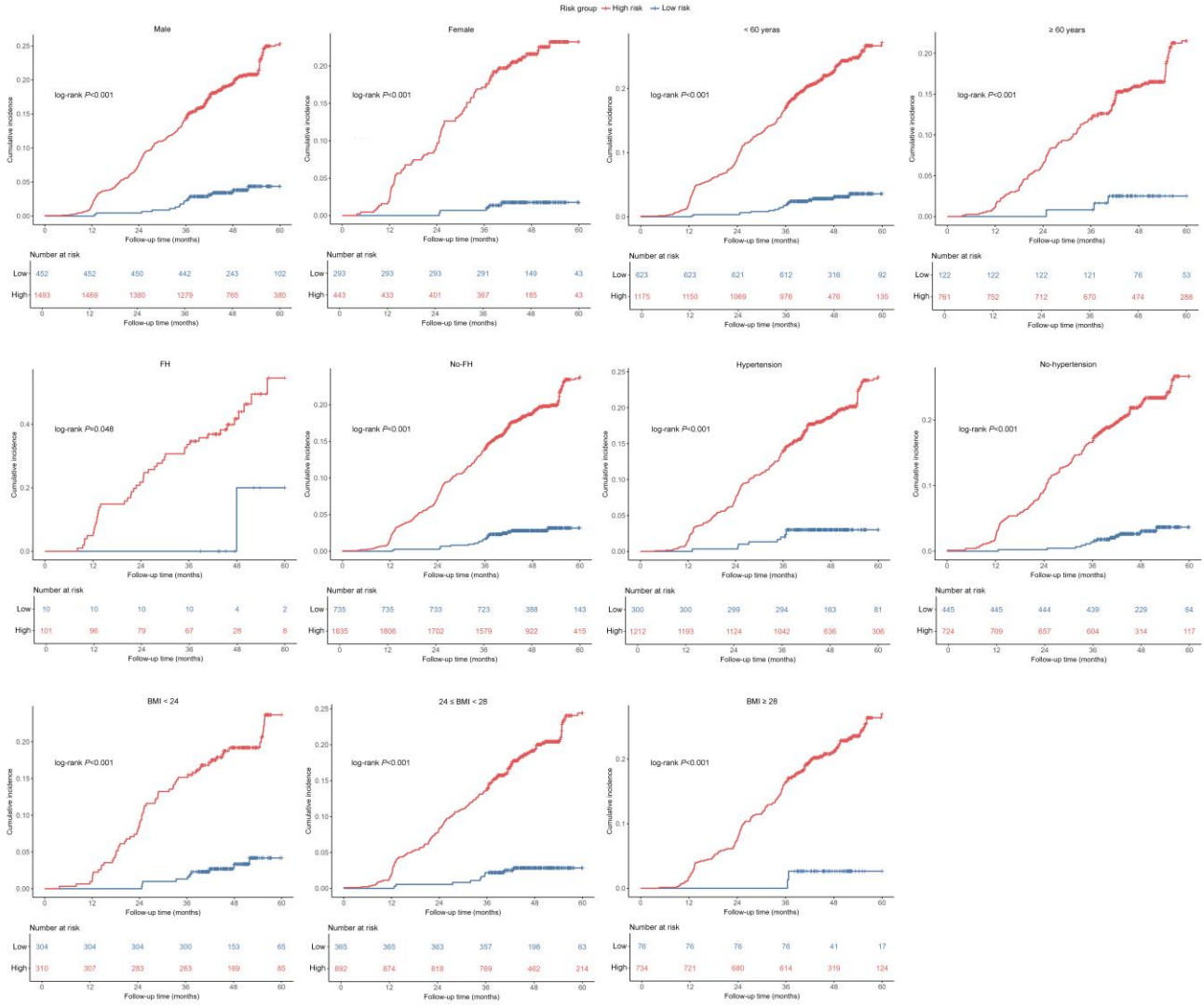
Supplementary Figure 2 The calibration curve in the external validation cohort.



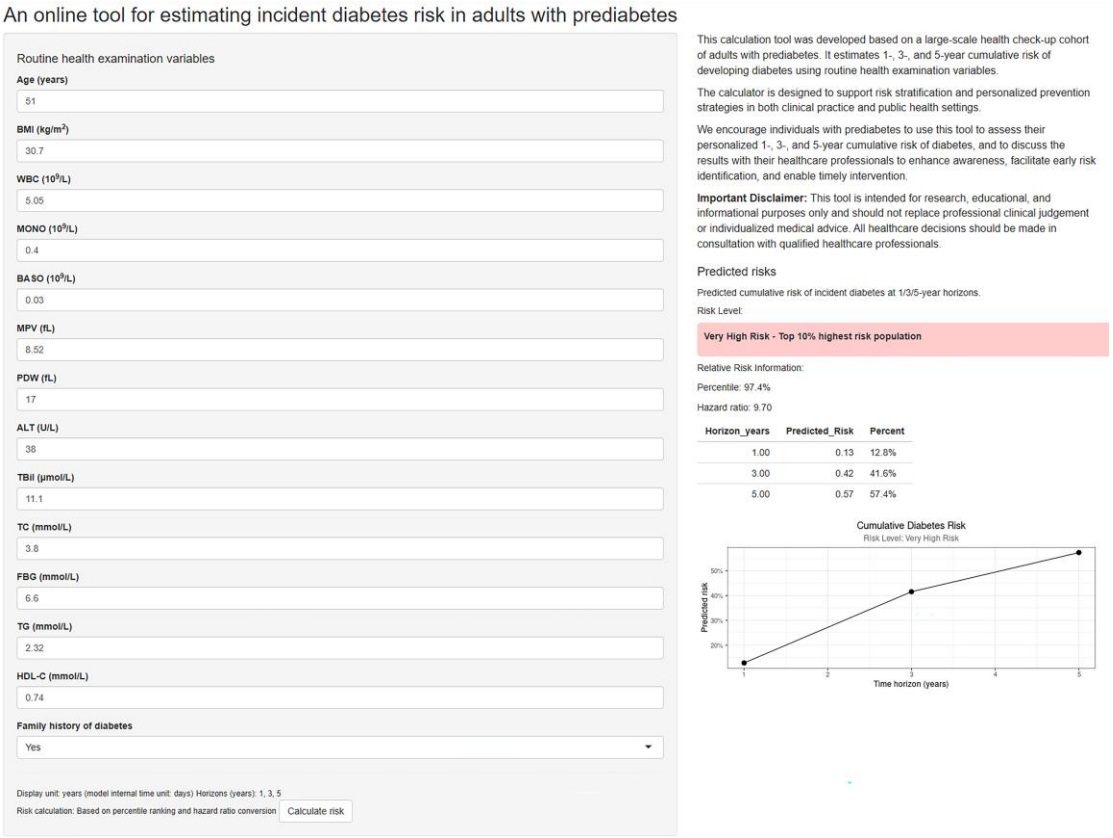
Supplementary Figure 3 Decision curve analysis of the gradient boosting survival analysis model and reference models at the 3-year prediction horizon. A: Testing cohort; B: External validation cohort. GBSA: Gradient boosting survival analysis; FBG: Fasting blood glucose.



Supplementary Figure 4 Kaplan-Meier curves comparing the cumulative incidence of type 2 diabetes mellitus in the high- and low-risk groups in the training cohort and testing cohort. A: Training cohort; B: Testing cohort.



Supplementary Figure 5 Subgroup analysis of the gradient boosting survival analysis model in the external validation cohort. Kaplan-Meier curves comparing the cumulative incidence of type 2 diabetes mellitus between the high-risk and low-risk groups across subgroups. GBSA: Gradient boosting survival analysis; FH: Family history; BMI: Body mass index.



Supplementary Figure 6 The web-based risk calculator for the gradient boosting survival analysis model. GBSA: Gradient boosting survival analysis; BMI: Body mass index; WBC: White blood cell count; MONO: Monocyte count; BASO: Basophil count; MPV: Mean platelet volume; PDW: Platelet distribution width; ALT: Alanine aminotransferase; TBil: Total bilirubin; TC: Total cholesterol; FBG: Fasting blood glucose; TG: Triglycerides; HDL-C: High-density lipoprotein cholesterol.

Supplementary Table 1 The proportions of missing data for retained candidate predictive variables before imputation across cohorts, *n* (%)

Variable	Training cohort (*n* = 15228)	Testing cohort (*n* = 9700)	Validation cohort (*n* = 2681)

Age	0 (0)	0 (0)	0 (0)
Sex	0 (0)	0 (0)	0 (0)
SBP	1235 (8.11)	209 (2.15)	184 (6.86)
DBP	1240 (8.14)	210 (2.16)	182 (6.79)
BMI	2647 (17.38)	441 (4.55)	160 (5.97)
WBC	508 (3.34)	6 (0.06)	151 (5.63)
Neutrophil count	164 (1.08)	6 (0.06)	151 (5.63)
Lymphocyte count	163 (1.07)	7 (0.07)	151 (5.63)
MONO	163 (1.07)	7 (0.07)	151 (5.63)
EOS	2800 (18.39)	7 (0.07)	151 (5.63)
BASO	2800 (18.39)	7 (0.07)	151 (5.63)
HGB	164 (1.08)	6 (0.06)	151 (5.63)
HCT	2800 (18.39)	7 (0.07)	151 (5.63)
MCV	222 (1.46)	7 (0.07)	151 (5.63)
MCH	2800 (18.39)	8 (0.08)	151 (5.63)
MCHC	2800 (18.39)	8 (0.08)	151 (5.63)
PLT	2800 (18.39)	8 (0.08)	151 (5.63)
PCT	177 (1.16)	29 (0.30)	151 (5.63)
MPV	173 (1.14)	29 (0.30)	151 (5.63)
PDW	173 (1.14)	29 (0.30)	154 (5.74)
RDW-CV	165 (1.08)	7 (0.07)	151 (5.63)
ALT	22 (0.14)	3 (0.03)	41 (1.53)
AST	20 (0.13)	3 (0.03)	47 (1.75)
TBil	4427 (29.07)	385 (3.97)	86 (3.21)
DBil	4427 (29.07)	385 (3.97)	86 (3.21)
Creatinine	11 (0.07)	2 (0.02)	127 (4.74)
Urea	8 (0.05)	5 (0.05)	127 (4.74)
Uric acid	8 (0.05)	1 (0.01)	127 (4.74)
FBG	3 (0.02)	1 (0.01)	0 (0)
TC	11 (0.07)	1 (0.01)	16 (0.60)

TG	31 (0.20)	9 (0.09)	596 (22.23)
HDL-C	488 (3.20)	46 (0.47)	592 (22.08)
LDL-C	489 (3.21)	41 (0.42)	140 (5.22)
FH of T2DM	1758 (11.54)	2029 (20.92)	378 (14.10)

SBP: Systolic blood pressure; DBP: Diastolic blood pressure; BMI: Body mass index; WBC: White blood cell count; MONO: Monocyte count; EOS: Eosinophil count; BASO: Basophil count; HGB: Hemoglobin; HCT: Hematocrit; MCV: Mean corpuscular volume; MCH: Mean corpuscular hemoglobin; MCHC: Mean corpuscular hemoglobin concentration; PLT: Platelets; PCT: Plateletcrit; MPV: Mean platelet volume; PDW: Platelet distribution width; RDW-CV: Red blood cell distribution width-coefficient; ALT: Alanine aminotransferase; AST: Aspartate aminotransferase; TBil: Total bilirubin; DBil: Direct bilirubin; FBG: Fasting blood glucose; TC: Total cholesterol; TG: Triglycerides; HDL-C: High-density lipoprotein cholesterol; LDL-C: Low-density lipoprotein cholesterol; FH: Family history; T2DM: Type 2 diabetes mellitus.

Supplementary Table 2 The numbers and proportions of participants who met each diagnostic criterion for baseline prediabetes and incident type 2 diabetes mellitus, *n* (%)

Criteria	**Training cohort**		**Testing cohort**		**Validation cohort**	
	Identified by criteria	**Missing**	**Identified by criteria**	**Missing**	**Identified by criteria**	**Missing**
Baseline prediabetes	*n* = 15228		*n* = 9700		*n* = 2681	
FBG (5.6–6.9 mmol/L)	14189 (93.18)	3 (0.02)	5976 (61.61)	1 (0.01)	2681 (100)	0 (0)
HbA1c (5.7%–6.4%)	1778 (11.68)	12247 (80.42)	5433 (56.01)	2615 (26.96)	0 (0)	2681 (100)
Incident T2DM	*n* = 2413		*n* = 914		*n* = 454	
FBG (≥ 7.0 mmol/L)	1743	8 (0.33)	551 (60.28)	1 (0.11)	446	0 (0)

	(72.23)				(98.24)	
HbA1c (≥ 6.5%)	638 (26.44)	1420 (58.85)	472 (51.64)	125 (13.68)	5 (1.10)	447 (98.46)
Hypoglycemic medication use	170 (7.05)	1882 (77.99)	0 (0)	654 (71.55)	8 (1.76)	90 (19.82)
Self-reported clinical diagnosis	313 (12.97)	1215 (50.35)	107 (11.71)	338 (36.98)	18 (3.96)	90 (19.82)

Because diagnostic criteria were not mutually exclusive, percentages may sum to more than 100%. FBG: Fasting blood glucose; HbA1c: Hemoglobin A1c; T2DM: Type 2 diabetes mellitus.

Supplementary Table 3 Baseline characteristics of study participants by progression status, mean ± SD/*n* (%)/median (Q1, Q3)

Variable	Training cohort			Testing cohort			Validation cohort		
	Progression (*n* = 2413)	Non-Progression (*n* = 12815)	*P* value	Progression (*n* = 914)	Non-Progression (*n* = 8786)	*P* value	Progression (*n* = 454)	Non-Progression (*n* = 2227)	*P* value
Age (years)	53.39 ± 12.69	47.87 ± 13.84	< 0.001	51.61 ± 12.89	45.90 ± 12.25	< 0.001	53.57 ± 12.12	52.32 ± 12.73	0.056
Sex			0.004			< 0.001			0.015
Male	1481 (61.38)	7454 (58.17)		636 (69.58)	5506 (62.67)		351 (77.31)	1594 (71.58)	
Female	932 (38.62)	5361 (41.83)		278 (30.42)	3280 (37.33)		103 (22.69)	633 (28.42)	
BMI (kg/m^2)	26.38 ± 3.54	24.77 ± 3.42	< 0.001	26.73 ± 3.76	25.04 ± 3.51	< 0.001	27.34 ± 3.44	26.35 ± 3.28	< 0.001
SBP (mmHg)	132.14 ± 17.48	126.99 ± 17.61	< 0.001	132.94 ± 17.39	126.74 ± 16.56	< 0.001	142.97 ± 19.21	142.11 ± 19.51	0.406
DBP (mmHg)	79.41 ± 11.72	76.98 ± 11.55	< 0.001	80.02 ± 11.79	76.19 ± 11.44	< 0.001	87.18 ± 11.89	86.08 ± 11.91	0.083
FBG (mmol/L,)	6.18 (5.88, 6.53)	5.81 (5.67, 6.03)	< 0.001	6.04 (5.72, 6.41)	5.64 (5.27, 5.82)	< 0.001	6.26 (5.96, 6.62)	5.83 (5.69, 6.07)	< 0.001
WBC (10^9/L)	6.19 (5.12, 7.30)	5.63 (4.50, 6.77)	< 0.001	6.46 (5.55, 7.56)	6.00 (5.14, 7.00)	< 0.001	6.30 (5.43, 7.40)	6.07 (5.22, 7.07)	< 0.001
Neutrophil count (10^9/L)	3.67 (3.00, 4.44)	3.44 (2.80, 4.20)	< 0.001	3.65 (3.00, 4.50)	3.36 (2.72, 4.10)	< 0.001	3.70 (3.00, 4.50)	3.50 (2.80, 4.20)	< 0.001
Lymphocyte	2.07 (1.71,	1.98 (1.63, 2.37)	< 0.001	2.10 (1.71, 2.52)	2.00 (1.68, 2.40)	< 0.001	1.90 (1.60,	1.90 (1.60, 2.30)	0.169

count (10^9/L)	2.51)						2.40)		
MONO (10^9/L)	0.38 (0.30, 0.48)	0.33 (0.27, 0.42)	< 0.001	0.43 (0.37, 0.53)	0.40 (0.32, 0.50)	< 0.001	0.47 (0.40, 0.58)	0.40 (0.40, 0.50)	0.017
EOS (10^9/L)	0.16 ± 0.13	0.14 ± 0.12	< 0.001	0.17 ± 0.14	0.15 ± 0.13	< 0.001	0.16 ± 0.20	0.13 ± 0.12	< 0.001
BASO (10^9/L)	0.03 ± 0.02	0.03 ± 0.02	0.165	0.03 ± 0.03	0.03 ± 0.03	< 0.001	0.01 ± 0.03	0.01 ± 0.03	0.266
PLT (10^9/L)	239.75 ± 59.89	238.45 ± 55.37	0.327	240.47 ± 61.17	244.11 ± 55.30	0.061	229.60 ± 59.46	227.87 ± 57.39	0.569
MPV (fL)	8.70 (8.00, 9.70)	8.60 (8.00, 9.50)	< 0.001	9.70 (8.90, 10.50)	9.70 (8.80, 10.50)	0.382	8.46 (7.75, 9.20)	8.40 (7.77, 9.14)	0.380
PDW (fL)	15.11 ± 2.06	15.35 ± 1.83	< 0.001	13.44 ± 2.71	13.49 ± 2.72	0.628	16.91 ± 1.08	16.85 ± 1.13	0.352
HGB (g/L)	147.37 ± 14.39	146.89 ± 14.91	0.140	148.08 ± 14.69	144.99 ± 15.18	< 0.001	146.36 ± 14.01	145.07 ± 14.53	0.087
TC (mmol/L)	5.39 ± 1.06	5.16 ± 0.90	< 0.001	6.06 ± 0.50	5.57 ± 0.44	< 0.001	4.88 ± 0.98	4.89 ± 0.93	0.962
TG (mmol/L)	1.60 (1.12, 2.29)	1.29 (0.90, 1.89)	< 0.001	1.62 (1.08, 2.35)	1.32 (0.92, 1.91)	< 0.001	1.75 (1.30, 2.54)	1.56 (1.14, 2.20)	< 0.001
HDL-C (mmol/L)	1.20 (1.03, 1.39)	1.30 (1.10, 1.56)	< 0.001	1.09 (0.94, 1.31)	1.19 (1.02, 1.42)	< 0.001	1.13 (0.96, 1.30)	1.22 (1.03, 1.45)	< 0.001
LDL-C (mmol/L)	3.00±0.81	2.96±0.76	0.014	3.11±0.86	3.08±0.80	0.435	2.65±0.68	2.62±0.67	0.448
ALT (U/L)	24.00 (17.00, 36.00)	20.00 (15.00, 30.00)	< 0.001	27.00 (19.00, 42.00)	23.00 (16.00, 34.00)	< 0.001	22.00 (16.00, 32.00)	20.00 (14.00, 28.00)	< 0.001

AST (U/L)	21.00 (18.00, 26.00)		19.00 (17.00, 24.00)		< 0.001	22.00 (18.00, 27.00)		21.00 (17.00, 25.00)		< 0.001	20.00 (16.00, 26.00)		20.00 (17.00, 24.00)		0.144
TBil (μmol/L)	12.70 (9.74, 16.24)		13.34 (10.37, 17.15)		< 0.001	11.82 (9.20, 15.75)		11.73 (9.19, 15.07)		0.252	15.25 (12.30, 19.00)		15.10 (12.00, 19.20)		0.621
FH of T2DM					< 0.001					< 0.001					< 0.001
Yes	433 (17.94)		1146 (8.94)			193 (21.12)		943 (10.73)			38 (8.37)		52 (2.33)		
No	1681 (69.66)		10210 (79.67)			534 (58.42)		6001 (68.30)			358 (78.85)		1855 (83.30)		
Missing	299 (12.39)		1459 (11.39)			187 (20.46)		1842 (20.97)			58 (12.78)		320 (14.37)		

SD: Standard deviation; BMI: Body mass index; SBP: Systolic blood pressure; DBP: Diastolic blood pressure; FBG: Fasting blood glucose; WBC: White blood cell count; MONO: Monocyte count; EOS: Eosinophil count; BASO: Basophil count; PLT: Platelets; MPV: Mean platelet volume; PDW: Platelet distribution width; HGB: Hemoglobin; TC: Total cholesterol; TG: Triglycerides; HDL-C: High-density lipoprotein cholesterol; LDL-C: Low-density lipoprotein cholesterol; ALT: Alanine aminotransferase; AST: Aspartate aminotransferase; TBil: Total bilirubin; FH: Family history; T2DM: Type 2 diabetes mellitus.

Supplementary Table 4 Variables associated with incident type 2 diabetes mellitus in univariate Cox regression analysis

Variable	HR	95% CI	*P* value
FBG	9.42	8.52-10.40	< 0.001
BMI	1.11	1.10-1.13	< 0.001
Age	1.03	1.02-1.03	< 0.001
MONO	7.86	6.11-10.10	< 0.001
TG	1.20	1.17-1.23	< 0.001
HDL-C	0.37	0.32-0.42	< 0.001
FH of T2DM	2.10	1.89-2.34	< 0.001
WBC	1.12	1.10-1.14	< 0.001
TC	1.30	1.25-1.35	< 0.001
SBP	1.01	1.01-1.02	< 0.001
ALT	1.01	1.01-1.01	< 0.001
Urea	1.12	1.09-1.15	< 0.001
Lymphocyte count	1.33	1.25-1.42	< 0.001
PDW	0.92	0.90-0.94	< 0.001
Neutrophil count	1.14	1.10-1.17	< 0.001
Uric acid	1.00	1.00-1.00	< 0.001
MPV	1.14	1.10-1.18	< 0.001
EOS	2.36	1.77-3.14	< 0.001
TBil	0.98	0.97-0.99	< 0.001
Sex	1.13	1.04-1.23	0.003
PLT	1.00	1.00-1.00	0.030
MCV	1.01	1.00-1.02	0.037

T2DM: Type 2 diabetes mellitus; HR: Hazard ratio; CI: Confidence interval; FBG: Fasting blood glucose; BMI: Body mass index; MONO: Monocyte count; TG: Triglycerides; HDL-C: High-density lipoprotein cholesterol; FH: Family history; WBC: White blood cell count; TC: Total cholesterol; SBP: Systolic blood pressure; ALT: Alanine aminotransferase; PDW: Platelet distribution width; MPV: Mean platelet

volume; EOS: Eosinophil count; TBil: Total bilirubin; PLT: Platelets; MCV: Mean corpuscular volume.

Supplementary Table 5 Variables retained by least absolute shrinkage and selection operator Cox regression with shrunken coefficients and hazard ratios at the chosen lambda

Variable	Shrunken coefficient	HR
FBG	1.84	6.28
BMI	0.04	1.04
Age	0.02	1.02
MONO	0.74	2.10
TG	0.04	1.04
HDL-C	-0.31	0.74
FH of T2DM	0.58	1.78
WBC	0.04	1.04
TC	0.02	1.02
ALT	< 0.01	1.00
PDW	-0.03	0.97
MPV	0.04	1.04
TBil	-0.01	0.99
BASO	-0.24	0.79

Variables were retained at lambda.1se = 0.0118. HR: Hazard ratio; FBG: Fasting blood glucose; BMI: Body mass index; MONO: Monocyte count; TG: Triglycerides; HDL-C: High-density lipoprotein cholesterol; FH: Family history; T2DM: Type 2 diabetes mellitus; WBC: White blood cell count; TC: Total cholesterol; ALT: Alanine aminotransferase; PDW: Platelet distribution width; MPV: Mean platelet volume; TBil: Total bilirubin; BASO: Basophil count.

Supplementary Table 6 Incremental predictive performance of the final gradient

boosting survival analysis model compared with reference models

Cohort	Comparison	Reference model C-index (95%CI)		ΔC-index	Continuous NRI	IDI
Testing cohort	GBSA *vs* FBG-only	0.767	(0.749–0.784)	0.046	0.250	0.071
	GBSA *vs* parsimonious clinical model	0.800	(0.785–0.815)	0.013	0.050	0.052
Validation cohort	GBSA *vs* FBG-only	0.757	(0.734–0.779)	0.002	0.286	0.042
	GBSA *vs* parsimonious clinical model	0.756	(0.735–0.778)	0.003	0.332	0.028

Net reclassification improvement and integrated discrimination improvement were calculated for 3-year predicted risk. GBSA: Gradient boosting survival analysis; FBG: Fasting blood glucose; C-index: Concordance index; CI: Confidence interval; NRI: Net reclassification improvement; IDI: Integrated discrimination improvement.

Supplementary Table 7 Baseline characteristics of the risk groups in the external validation cohort, mean ± SD/*n* (%)/median (Q1, Q3)

Variable	High risk (*n* = 1936)	Low risk (*n* = 745)	*P* value
Age (years)	55.12 ± 11.61	45.82 ± 12.70	< 0.001
Sex			< 0.001
Male	1493 (77.12)	452 (60.67)	
Female	443 (22.88)	293 (39.33)	
BMI (kg/m^2)	27.24 ± 3.28	24.57 ± 2.62	< 0.001
SBP (mmHg)	144.86 ± 19.08	135.15 ± 18.71	< 0.001
DBP (mmHg)	87.43 ± 11.62	83.07 ± 12.10	< 0.001
FBG (mmol/L)	6.01 (5.81, 6.30)	5.70 (5.64, 5.78)	< 0.001
WBC (10^9/L)	6.29 (5.40, 7.27)	5.70 (4.89, 6.61)	< 0.001

Neutrophil count (10^9/L)	3.60 (3.00, 4.40)	3.20 (2.60, 3.80)	< 0.001
Lymphocyte count (10^9/L)	1.93 (1.60, 2.38)	1.90 (1.50, 2.30)	0.001
MONO (10^9/L)	0.50 (0.40, 0.59)	0.40 (0.30, 0.50)	< 0.001
EOS (10^9/L)	0.14 ± 0.15	0.12 ± 0.11	< 0.001
BASO (10^9/L)	0.01 ± 0.03	0.01 ± 0.03	0.645
PLT (10^9/L)	226.57 ± 58.64	232.40 ± 55.13	0.023
MPV (fL)	8.40 (7.78, 9.17)	8.40 (7.75, 9.12)	0.431
PDW (fL)	16.89 ± 1.18	16.79 ± 0.94	0.043
HGB (g/L)	146.65 ± 13.31	141.68 ± 16.57	< 0.001
TC (mmol/L)	4.91 ± 0.98	4.82 ± 0.80	0.018
TG (mmol/L)	1.73 (1.27, 2.48)	1.30 (0.98, 1.76)	< 0.001
HDL-C (mmol/L)	1.15 (0.98, 1.35)	1.32 (1.15, 1.56)	< 0.001
LDL-C (mmol/L)	2.65 ± 0.70	2.58 ± 0.59	0.022
ALT (U/L)	21.00 (16.00, 30.00)	17.00 (13.00, 25.00)	< 0.001
AST (U/L)	20.00 (17.00, 25.00)	19.00 (16.00, 23.00)	< 0.001
TBil (μmol/L)	15.25 (12.10, 19.40)	15.00 (12.00, 18.90)	0.335
FH of T2DM			< 0.001
Yes	84 (4.34)	6 (0.81)	
No	1556 (80.37)	657 (88.19)	
Missing	296 (15.29)	82 (11.01)	

SD: Standard deviation; BMI: Body mass index; SBP: Systolic blood pressure; DBP: Diastolic blood pressure; FBG: Fasting blood glucose; WBC: White blood cell count; MONO: Monocyte count; EOS: Eosinophil count; BASO: Basophil count; PLT: Platelets; MPV: Mean platelet volume; PDW: Platelet distribution width; HGB: Hemoglobin; TC: Total cholesterol; TG: Triglycerides; HDL-C: High-density lipoprotein cholesterol; LDL-C: Low-density lipoprotein cholesterol; ALT: Alanine aminotransferase; AST: Aspartate aminotransferase; TBil: Total bilirubin; FH: Family

history; T2DM: Type 2 diabetes mellitus.

Supplementary Table 8 Numbers of participants at risk, outcome events, and censored observations at 1-year, 3-year, and 5-year prediction horizons in the external validation cohort (*n* = 2681), *n* (%)

Prediction horizon	Participants at risk	Cumulative outcome events	Censored before horizon
1-year	2647 (98.73)	34 (1.27)	0 (0)
3-year	2379 (88.74)	302 (11.26)	0 (0)
5-year	568 (21.19)	437 (16.30)	1676 (62.51)