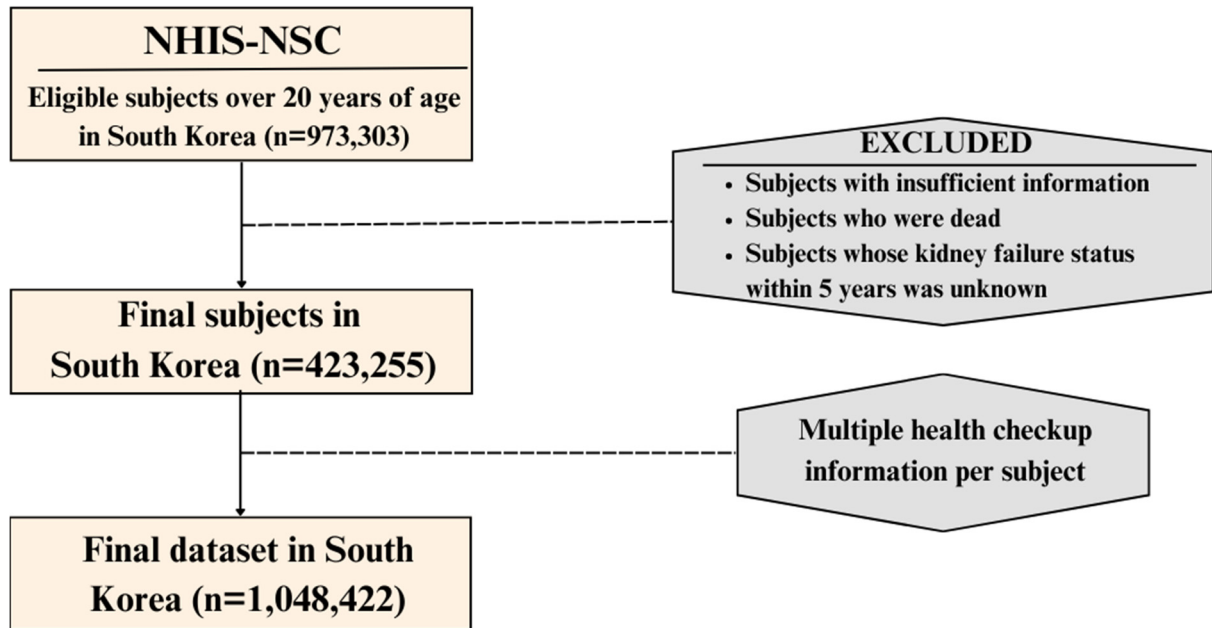


Supplementary Material

Figure S1. Data selection process in NHIS-NSC. Abbreviation: NHIS-NSC, National Health Insurance Service-National Sample Cohort.



7 **Table S1.** Chronic kidney disease-related diagnosis codes in ICD-10.

	ICD-10 codes
Chronic kidney disease	N170, N171, N172, N178, N179, N180, N189, N19, N990
Tubulo-interstitial nephritis	N110, N111, N118, N119, N12, N140, N141, N142, N143, N144, N150
Chronic glomerular nephritis	N002, N003, N004, N006, N007, N009, N012, N014, N016, N17, N019, N028, N029, N030, N032, N033, N034, N036, N037, N039, N040, N042, N044, N046, N049, N050, N051, N052, N053, N054, N055, N056, N057, N058, N059, N069, N079, N085
Diabetic nephropathy	E102, E112, E132, E142
Hypertensive nephrosclerosis	I129, I120
Polycystic kidney disease	Q613, Q612

8 Abbreviation: ICD-10, International Classification of Diseases, 10th revision.

9

10 **Table S2.** Included features for the machine learning model.

Value, n	Variables	Type	Description
1	Age	Numerical	Numerical
2	Sex	2 categories	<ul style="list-style-type: none"> • 0: male • 1: female
3	Household income	11 categories	<ul style="list-style-type: none"> • 0: basic livelihood recipient • 1: 0-9 percentile • 2: 10-19 percentile • 3: 20-29 percentile • 4: 30-39 percentile • 5: 40-49 percentile • 6: 50-59 percentile • 7: 60-69 percentile • 8: 70-79 percentile • 9: 80-89 percentile • 10: 90-100 percentile
4	Region of residence	2 categories	<ul style="list-style-type: none"> • 0: rural • 1: urban
5	Body mass index	Numerical	kg/m ²
6	Systolic blood pressure	Numerical	mmHg
7	Diastolic blood pressure	Numerical	mmHg
8	Fasting blood glucose	Numerical	mg/dL
9	Serum total cholesterol	Numerical	mg/dL
10	Hemoglobin	Numerical	g/dL
11	Aspartate transaminase	Numerical	U/L
12	Alanine transaminase	Numerical	U/L
13	γ -glutamyl transpeptidase	Numerical	U/L
14	history of chronic kidney disease related diagnosis	2 categories	<ul style="list-style-type: none"> • 0: no • 1: yes
15	History of diabetes mellitus	2 categories	<ul style="list-style-type: none"> • 0: no • 1: yes
16	History of stroke	2 categories	<ul style="list-style-type: none"> • 0: no • 1: yes
17	History of hypertension	2 categories	<ul style="list-style-type: none"> • 0: no • 1: yes
18	Smoking status	3 categories	<ul style="list-style-type: none"> • 1: never smoker • 2: ex-smoker

			<ul style="list-style-type: none"> • 3: current smoker
19	Alcoholic drinks	4 categories	<ul style="list-style-type: none"> • 1: <1 day per week • 2: 1-2 days per week • 3: 3-4 days per week • 4: ≥ 5 days per week
20	Physical activity sessions	5 categories	<ul style="list-style-type: none"> • 1: 0 day per week • 2: 1-2 days per week • 3: 3-4 days per week • 4: 5-6 days per week • 5: 7 days per week

11

12

13 **Table S3.** Summary of training and testing datasets.

	Kidney failure group	Non-kidney failure group	Total
Training data, n	10,509	828,053	838,562
Testing data, n	2,647	207,213	209,860
Total, n	13,156	1,035,266	1,048,422

14

15 **Table S4.** Comparison of the five-fold cross-validation results of model of our method, model
 16 trained on all features and model trained on features without age.

Models	Matrix, mean (SD)				
	Accuracy	Specificity	Sensitivity	Balanced accuracy	AUROC
Two model ensemble approach (The final model)	0.6962	0.6964	0.6861	0.6912	0.7541
	(0.0031)	(0.0033)	(0.0182)	(0.0086)	(0.0076)
All features	0.6823	0.6820	0.6996	0.6908	0.7530
	(0.0038)	(0.0039)	(0.0164)	(0.0081)	(0.0080)
Features without age	0.7041	0.7050	0.6356	0.6703	0.7286
	(0.0041)	(0.0043)	(0.0120)	(0.0043)	(0.0056)

17 Abbreviation: AUROC, area under receiver of characteristics.

18

19 **Table S5.** Ranked normalized feature importance values.

Rank	Name	Value
1	Age	0.19744
2	Body mass index	0.12383
3	Fasting blood glucose	0.11914
4	Diastolic blood pressure	0.08266
5	Systolic blood pressure	0.07565
6	Hemoglobin	0.07156
7	Physical activity sessions	0.05530
8	Alcoholic drinks	0.05384
9	History of stroke	0.04346
10	Aspartate transaminase	0.03489
11	Smoking state	0.03273
12	γ -glutamyl transpeptidase	0.03089
13	Alanine transaminase	0.02796
14	history of chronic kidney disease related diagnosis	0.02459
15	Serum total cholesterol	0.01203
16	History of diabetes mellitus	0.00561
17	Income	0.00428
18	History of hypertension	0.00217
19	Sex	0.00153
20	Region	0.00043

20