1 Supplementary data

2 Table S1. Correspondence table in routine urine characteristic values to numerical valence numbers.

Project	Neg	+/-	1+	2+	3+	4+
Glucose	0	30	50	150	500	1000
Protein	0	15	30	100	500	-
Bilirubin	0	0.5	1	3	6	-
Urobilinogen	0	-	2	4	8	12
Occult blood	0	-	5~10	50	300	-
Ketone body	0	5	15	50	150	-
Nitrite	0	-	Pos	-	-	-
Leukocytes	0	-	25	75	500	-
Bacteria	0	270	758	1237	4796	-
Bacteria	0	270	/58	1237	4796	

3 Non-numeric fields are represented using this format for conversion.

4 Neg: Negative, were denoted as 0

5 -: indicates no	specific	grading
-------------------	----------	---------

6

7

8

9 Table S2. Results for generating noise at various standard deviations using Gaussian noise. The

10 maximum discrimination ability on Validation AUC is only 0.79, which is 0.04 lower than that of the

11 original RF model (Validation AUC = 0.83).

	Valida	Validation					Independent		
	SN	SP	Precision	F1-score	ACC	AUC	ACC	AUC	
SD = 0.01	0.78	0.70	0.71	0.74	0.74	0.79	0.68	0.70	
SD = 0.1	0.70	0.72	0.70	0.70	0.71	0.75	0.63	0.61	
SD = 0.15	0.70	0.52	0.58	0.63	0.61	0.67	0.62	0.65	

12

13

14

16 Table S3. Performance Metrics Across Different Feature Screening Met	hods.
---	-------

	Validation			Independent				
	SN	SP	Precision	F1-score	ACC	AUC	ACC	AUC
num in RF	0.74	0.78	0.80	0.74	0.76	0.83	0.57	0.64
4×4_black_num_full_features ^a	0.80	0.68	0.75	0.80	0.75	0.80	0.63	0.69
4×4_black_num_mrmr(MIQ) ^b	0.79	0.68	0.74	0.79	0.74	0.80	0.64	0.69
4×4_black_num_pca ^b	0.80	0.65	0.73	0.80	0.73	0.81	0.59	0.61
4×4_black_num_svd ^b	0.80	0.68	0.75	0.80	0.75	0.82	0.61	0.53
4×4_black_num_tsne ^c	0.84	0.71	0.77	0.84	0.78	0.83	0.58	0.61
4×4 _uniform_noise_full_features ^d	0.78	0.75	0.79	0.78	0.76	0.80	0.66	0.69
4×4_uniform_noise_mrmr(MIQ) ^b	0.80	0.68	0.75	0.80	0.75	0.81	0.64	0.69
4×4_uniform_noise_pca ^b	0.75	0.71	0.75	0.75	0.73	0.80	0.45	0.52
4×4 _uniform_noise_svd ^b	0.78	0.69	0.75	0.78	0.74	0.80	0.65	0.52
4×4_uniform_noise_tsne ^c	0.82	0.72	0.78	0.82	0.78	0.83	0.54	0.64
combine_uniform_noise_full_featurese	0.71	0.61	0.68	0.71	0.66	0.78	0.63	0.68
combine_uniform_noise_mrmr(MIQ) ^b	0.74	0.71	0.75	0.74	0.73	0.82	0.53	0.64
combine_uniform_noise_pcab	0.80	0.67	0.74	0.80	0.74	0.80	0.55	0.56
combine_uniform_noise_svd ^b	0.79	0.65	0.73	0.79	0.73	0.80	0.52	0.64
combine_uniform_noise_tsnec	0.79	0.72	0.77	0.79	0.76	0.83	0.47	0.50

17 ^{a.} Extracting 24 feature vectors; ^{b.} Extracting 13 feature vectors; ^{c.} Extracting 2 feature vectors; ^{d.} Extracting 20

18 feature vectors; ^{e.} Extracting 64 feature vectors

19

- 21 Table S4. Comparison of demographic and clinical characteristics between the Fengyuan and MacKay
- 22 datasets.

Fasture (Demographies)	Equation $(n-793)$	MaaKay (n-106)		
Candar (Mala 9/)	<u>rengyuan (n-765)</u>	122 (67 259()		
Gender (Male, $\%$)	434(57.98%)	132(07.35%) 67.01 ± 0.56		
Age (Mean \pm SD)	60.47 ± 13.21	67.01 ± 9.30		
pH (Mean \pm SD)	6.03 ± 0.91	6.17 ± 0.91		
Specific Gravity (S.G.)	1.013 (1.001, 1.039)	1.015 (1.002, 1.043)		
(Median, IQR)				
Feature (Urinalysis				
Biomarkers)				
Giucose (%)	600(76620/)	154 (78 570/)		
neg	2(0.26%)	134(78.3770) 12(6620/)		
± 1	2(0.2070)	13(0.05%) 12(6.629/)		
1+ 2+	33 (11.2470)	(0.05%)		
2+	54(4.54%)	0(5.00%)		
	39 (7.34%)	10 (3.10%)		
Protein (%)	451 (57 600/)	06 (48 080/)		
neg	431(37.00%) 121(16729/)	90(48.9870) 22(16.940/)		
1+	131(10.75%)	33(10.84%)		
2+	101(20.30%)	32(10.33%) 25(17.86%)		
5+ Dilimbir (0/)	40 (3.11%)	33 (17.80%)		
Billrubin (%)	777 (00 220/)	192 (02 9(0/)		
Neg	/// (99.23%) 5 (0.649()	182 (92.86%)		
1+	5(0.64%)	13(0.03%)		
<u></u>	1 (0.13%)	1 (0.51%)		
Urobiinogen (%)	720 (02 100/)	162 (82 160/)		
Neg	/29 (93.10%)	163 (83.16%)		
1+	33(4.21%)	34 (17.35%)		
	21 (2.68%)	0 (0%)		
Occult blood (%)	295 (26 409/)	94 (42 960/)		
Neg	285 (36.40%)	84 (42.86%)		
±	0(0%)	22(11.22%)		
1+	194 (24.78%)	19 (9.69%)		
2+	123(15./1%)	1/(8.07%)		
S⊤ Veterse hedre (0/)	181 (23.12%)	34 (27.33%)		
Ketone body (%)	722 (02 210/)	177 (00 219/)		
neg	/22 (92.21%)	1/7(90.5176) 15(7(597))		
±	47(0.00%)	13(7.03%)		
1+ 2+	10(1.2870) 4(0.5194)	4(2.0470)		
<u> </u>	+ (0.3170)	0 (070)		
Nitt tte (70)	732 (03 400/)	170 (86 720/)		
	(32 (33.4970) 51 (6 510/)	1/0(00./370) 26(12,270/)		
L outros (9/)	51 (0.5170)	20 (13.2776)		
Leukocytes (%)	406 (62 259/)	05 (48 479/)		
neg	490 (03.33%)	93 (48.4770) 21 (10 719/)		
± 1+	0(070) 102(12159/)	21(10.7170) 26(12.2794)		
1 F 2+	79 (10 00%)	20(13.2770) 20(10,200%)		
∠ Γ 3+	105 (13 /10/2)	20(10.2070) 34(17350)		
Sausmours (0/.)	103 (13.4170)	54 (17.5570)		
54uamous (70)	754 (06 200/)	133 (67 960/)		
$0 \sim 3$	15 (1 02%)	155 (07.8070)		
0~10 10~20	7(0.80%)	26(12,270)		
10~20 20- 20	/ (0.0970) 2 (0.290/)	20(13.2770) 8(1080/)		
20~30	2(0.3870)	o (4.0070) 5 (2 55%)		
50~50	2(0.2070) 2(0.2604)	S (2.3370) S (1 000/)		
50~100	2 (0.2070)	0 (4.0070)		



Figure S1. Interface of the web-based prediction tool. This platform allows patients to input routine urine detection values, which are then processed by the underlying algorithm to deliver predictive outcomes. The main interface comprises an input panel on the left side, where patients can manually enter values for various parameters such as age, gender, glucose, protein, and other urine-related indicators. After inputting/selecting the necessary data, by clicking the "Submit" button situated below the input fields, the tool processes the data, and a prediction result is displayed on the right side of the screen.



4x4_black_num

Figure S2. The model stability of numerical data on a 4×4 black background. When the initial filters
 range from 6 to 8, the model demonstrates enhanced identification ability, with performance closely

36 approaching 0.75 in terms of AUC during validation. Notably, the model exhibits the highest stability



37 when the parameter is set to 7.

Figure S3. The model stability of binary data on an 8×8 black background. When the number of initial
channels ranges from 3 to 16, the model exhibits higher recognition ability, with all values approaching
0.75. However, compared to the numerical data on a 4×4 black background, the stability of the model
is relatively poorer in this case.





45 screening to identify those with high stability for further analysis. In the performance of replicating and





16x16_black_num

Figure S5. The model stability of flipping and replicating a 16×16 image. The results are similar to
those obtained when replicating and flipping a 16×16 image, with the recognition ability still lower
than that without flipping.



8x8_black_num

51

Figure S6. The model stability of simple replication at 8×8 image. Compared to the performance of
16×16, the results of simple replication at 8×8 demonstrate higher recognition ability. Among these

54 results, parameter 8 exhibits the highest recognition ability.



55

56 Figure S7. The model stability of uniform noise at a maximum value of 5. Compared to replicating and

57 flipping methods, uniform noise exhibits higher recognition ability and stability. When the parameter

58 ranges from 5 to 8. The medians for all these parameters are close to 0.75.



max(10)

59

60 Figure S8. The model stability of uniform noise at a maximum value of 10. Compared to uniform noise

61 at a maximum value of 5, the stability decreases when the parameter ranges from 6 to 8. However,

62 when the parameter is set to 5, both stability and identification ability are higher.



Figure S9. The model stability of uniform noise at a maximum value of 20. The stability deteriorates at
a parameter value of 5, and parameter 6 exhibits high stability but with more outliers. The overall
identification ability of parameters ranging from 5 to 16 is relatively good, with the median values
being close to 0.75.



max(40)

68

Figure S10. The model stability of uniform noise at a maximum value of 40. Parameters 5 to 16
demonstrate balanced stability, with parameter 8 showing slightly higher identification ability in terms

71 of the median.



Figure S11. The model stability of uniform noise at a maximum value of 80. Parameter 7 shows high

real stability but with outliers, while parameters 8 and 16 exhibit poor stability. The discrimination ability





max(160)

Figure S12. The model stability of uniform noise at a maximum value of 160. Parameter 5 has more
outliers, and the other parameters exhibit a slight decline in discrimination ability with varying uniform
noise brightness levels.

80



Figure S13. The model stability of uniform noise at a maximum value of 255. The difference among
parameters 5 to 16 is not substantial, with parameter 7 demonstrating higher stability and better
identification ability.





86 Figure S14. 4×4_black_num_model visualization results of SVD. The classification effect is not

87 obvious.



90 Figure S15. 4×4_uniform_noise model visualization results of SVD. The classification results are not





92

93 Figure S16. Combine uniform noise model visualization results of SVD. The classification results

94 exhibit a high degree of overlap.