

Supplemental Material

Advancing Depression Risk Assessment in Connective Tissue Diseases: A Multi-Classification Machine Learning Approach

Leilei Yang^{1*}, Yuzhan Jin^{2,3*}, Wei Lu^{2,3*}, Xiaoqin Wang¹, Yuqing Yan^{2,3}, Yulan Tong^{2,3}, Dinglei Su^{1#}, Kaizong Huang^{3,4#}, Jianjun Zou^{3,4#}

Supplementary Figures and Tables

Table S1	<i>Page 2-4</i>
Table S2	<i>Page 5</i>
Table S3	<i>Page 6</i>
Table S4	<i>Page 7</i>
Figure S1	<i>Page 8</i>
Figure S2	<i>Page 8</i>
Figure S3	<i>Page 9</i>

1 Supplemental Tables

Table S1 The difference of demographic, clinical and laboratory characteristics and outcome between the training and test cohorts

	Total (n = 480)	Training cohort (n = 384)	Test cohort (n = 96)	P
Age (year)	58.5 [48.8, 68.0]	58.0 [49.0, 68.0]	60.0 [47.0, 69.0]	0.731
Disease duration (year)	4.0 [0.5, 10.0]	4.0 [0.5, 10.0]	3.0 [0.9, 10.0]	0.463
Education time (year)	9.0 [6.0, 12.0]	9.0 [6.0, 12.0]	9.0 [6.0, 9.0]	0.027
Sleep duration (hour)	6.0 [5.0, 7.0]	6.0 [5.0, 7.0]	7.0 [6.0, 7.0]	0.002
Gender				0.172
Male	58 (12.1%)	42 (10.9%)	16 (16.7%)	0.834
Female	422 (87.9%)	342 (89.1%)	80 (83.3%)	
Marital status				
Married	413 (86.0%)	331 (86.2%)	82 (85.4%)	
Unmarried	65 (13.5%)	51 (13.3%)	14 (14.6%)	0.517
Divorced or widowed	2 (0.42%)	2 (0.52%)	0 (0.00%)	
Comorbidity				
Other CTDs				
Yes	99 (20.6%)	82 (21.4%)	17 (17.7%)	0.544
No	381 (79.4%)	302 (78.6%)	79 (82.3%)	
Thyroid disease				
Yes	93 (19.4%)	77 (20.1%)	16 (16.7%)	
No	387 (80.6%)	307 (79.9%)	80 (83.3%)	0.414
Diabetes				
Yes	56 (11.7%)	42 (10.9%)	14 (14.6%)	
No	424 (88.3%)	342 (89.1%)	82 (85.4%)	
Other chronic diseases				0.049
Yes	201 (42.1%)	152 (39.8%)	49 (51.6%)	0.767
No	276 (57.9%)	230 (60.2%)	46 (48.4%)	
Infection				
Yes	113 (23.5%)	92 (24.0%)	21 (21.9%)	
No	367 (76.5%)	292 (76.0%)	75 (78.1%)	1.000
Fibromyalgia				
Yes	18 (3.75%)	15 (3.91%)	3 (3.12%)	
No	462 (96.2%)	369 (96.1%)	93 (96.9%)	
Pain (VAS score)	1.99 (2.41)	1.94 (2.39)	2.16 (2.51)	0.453
Fatigue				0.084
None	172 (35.8%)	127 (33.1%)	45 (46.9%)	0.266
Less than 7 days in two weeks	174 (36.2%)	147 (38.3%)	27 (28.1%)	
More than 7 days in two weeks	93 (19.4%)	76 (19.8%)	17 (17.7%)	
Almost every day	41 (8.54%)	34 (8.85%)	7 (7.29%)	
DAS28-CRP	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]	0.0 [0.0, 1.8]	0.711
SLEDAI-2000	5.0 [3.0, 10.0]	5.0 [3.0, 8.5]	6.0 [3.0, 12.0]	
SLEDAS	8.57 [2.09, 19.1]	8.32 [2.08, 19.1]	11.1 [2.95, 17.3]	

Prednisone (mg/d)	5.00 [0.00, 25.0]	5.00 [0.00, 30.0]	5.00 [0.00, 23.0]	0.464
Biological agents				0.478
Yes	39 (8.12%)	29 (7.55%)	10 (10.4%)	
No	441 (91.9%)	355 (92.4%)	86 (89.6%)	
Classic immune preparations				0.465
Yes	90 (18.8%)	69 (18.0%)	21 (21.9%)	
No	390 (81.2%)	315 (82.0%)	75 (78.1%)	
Traditional anti-rheumatic drugs				1.000
Yes	282 (58.8%)	226 (58.9%)	56 (58.3%)	
No	198 (41.2%)	158 (41.1%)	40 (41.7%)	
WBC count (10 ⁹ /L)	5.85 [4.47, 7.90]	5.74 [4.37, 7.90]	6.17 [4.66, 7.94]	0.286
Hemoglobin (g/L)	117 [104, 129]	116 [103, 127]	122 [106, 132]	0.071
Platelet (10 ⁹ /L)	193 [140, 254]	190 [134, 254]	207 [156, 254]	0.193
Neutrophil count (10 ⁹ /L)	3.79 [2.59, 5.89]	3.76 [2.50, 5.83]	3.84 [2.95, 6.02]	0.450
Lymphocyte count (10 ⁹ /L)	1.21 [0.88, 1.65]	1.18 [0.87, 1.61]	1.29 [1.00, 1.76]	0.056
Anti-Ro52				0.261
Positive	217 (45.2%)	179 (46.6%)	38 (39.6%)	
Negative	263 (54.8%)	205 (53.4%)	58 (60.4%)	
APS				0.221
Positive	31 (6.51%)	28 (7.33%)	3 (3.19%)	
Negative	445 (93.5%)	354 (92.7%)	91 (96.8%)	
Anti-Rib-P				0.786
Positive	46 (9.58%)	38 (9.90%)	8 (8.33%)	
Negative	434 (90.4%)	346 (90.1%)	88 (91.7%)	
Anti-SSA				0.219
Positive	163 (34.0%)	136 (35.4%)	27 (28.1%)	
Negative	317 (66.0%)	248 (64.6%)	69 (71.9%)	
Anti-SSB				0.673
Positive	47 (9.79%)	36 (9.38%)	11 (11.5%)	
Negative	433 (90.2%)	348 (90.6%)	85 (88.5%)	
Anti-RNP/Sm				0.310
Positive	68 (14.2%)	58 (15.1%)	10 (10.4%)	
Negative	412 (85.8%)	326 (84.9%)	86 (89.6%)	
Anti-CENP-B				1.000
Positive	55 (11.5%)	44 (11.5%)	11 (11.5%)	
Negative	425 (88.5%)	340 (88.5%)	85 (88.5%)	
ILD				0.884
Yes	115 (24.0%)	93 (24.3%)	22 (22.9%)	
No	364 (76.0%)	290 (75.7%)	74 (77.1%)	
PAH				0.478
Yes	63 (13.1%)	53 (13.8%)	10 (10.4%)	
No	417 (86.9%)	331 (86.2%)	86 (89.6%)	
IL6	6.26 [3.60, 15.6]	6.37 [3.62, 16.1]	5.16 [3.60, 13.0]	0.177
ESR	45.0 [20.0, 85.0]	44.0 [19.0, 86.0]	45.0 [22.0, 80.0]	0.919
CRP	4.79 [2.31, 18.3]	4.86 [2.41, 19.2]	4.39 [1.88, 16.7]	0.302

C3	0.78 [0.65, 0.93]	0.78 [0.65, 0.93]	0.78 [0.66, 0.90]	0.923
C4	0.19 [0.15, 0.24]	0.19 [0.15, 0.24]	0.19 [0.14, 0.23]	0.677
IgG	13.9 [11.1, 17.6]	14.0 [11.2, 17.7]	13.5 [10.9, 16.4]	0.253
Depression				0.143
Normal	242 (50.4%)	185 (48.2%)	57 (59.4%)	
Mild	144 (30.0%)	121 (31.5%)	23 (24.0%)	
Moderate and severe	94 (19.6%)	78 (20.3%)	16 (16.7%)	

Notes: ML, machine learning; LR, logistic regression; SVM, support vector machine; LGBM, light gradient boosting machine; Catboost, categorical boosting; ANN, artificial neural network.

Table S2 Missing situation for the variables with missing values in the original dataset

Variables	Missing values count	Missing values rate (%)
DAS28-CRP	100	20.8
Prednisone	22	4.58
Sleep duration	1	0.208
Other chronic diseases	3	0.625
T-SPOT	237	49.4
WBC count	5	1.04
Hemoglobin	5	1.04
Platelet	5	1.04
Neutrophil count	5	1.04
Lymphocyte count	5	1.04
APS	4	0.83
IL2	142	29.6
IL10	142	29.6
IL4	142	29.6
TNF	142	29.6
IFNr	141	29.4
IL6	115	24.0
ESR	43	8.96
CRP	14	2.92
C3	10	2.08
C4	11	2.29
IgG	15	3.13
ILD	1	0.208

Table S3 The performance of the six multi-classification ML models with 10-fold cross validation in the training cohort (n = 384)

	Acc	Kappa	W_prec	W_recall	W_F1	None				Mild				Moderate and severe			
						AUC	Prec	Recall	F1	AUC	Prec	Recall	F1	AUC	Prec	Recall	F1
LR																	
Fold 1	0.667	0.458	0.663	0.667	0.662	0.880	0.767	0.831	0.798	0.753	0.518	0.532	0.525	0.840	0.642	0.486	0.553
Fold 2	0.672	0.467	0.671	0.672	0.668	0.885	0.761	0.825	0.792	0.751	0.531	0.550	0.541	0.863	0.673	0.500	0.574
Fold 3	0.667	0.455	0.664	0.667	0.659	0.890	0.769	0.843	0.805	0.750	0.517	0.560	0.537	0.842	0.644	0.414	0.504
Fold 4	0.661	0.447	0.654	0.661	0.653	0.886	0.773	0.843	0.807	0.758	0.517	0.550	0.533	0.835	0.583	0.400	0.475
Fold 5	0.671	0.463	0.665	0.671	0.665	0.884	0.769	0.838	0.802	0.755	0.532	0.541	0.536	0.852	0.623	0.471	0.537
Fold 6	0.633	0.399	0.625	0.633	0.625	0.871	0.723	0.814	0.766	0.742	0.486	0.477	0.481	0.849	0.608	0.443	0.512
Fold 7	0.662	0.447	0.657	0.662	0.655	0.892	0.761	0.838	0.798	0.760	0.518	0.541	0.529	0.845	0.625	0.429	0.508
Fold 8	0.650	0.428	0.643	0.650	0.643	0.875	0.742	0.826	0.782	0.745	0.509	0.505	0.507	0.838	0.615	0.457	0.525
Fold 9	0.647	0.423	0.640	0.647	0.640	0.875	0.738	0.826	0.780	0.741	0.509	0.509	0.509	0.835	0.608	0.437	0.508
Fold10	0.653	0.435	0.645	0.653	0.646	0.884	0.755	0.837	0.794	0.753	0.509	0.514	0.511	0.848	0.596	0.437	0.504
Mean	0.658	0.442	0.653	0.658	0.652	0.882	0.756	0.832	0.792	0.751	0.515	0.528	0.521	0.845	0.622	0.447	0.520
SVM																	
Fold 1	0.699	0.511	0.708	0.699	0.696	0.898	0.802	0.831	0.817	0.799	0.547	0.642	0.591	0.869	0.733	0.471	0.574
Fold 2	0.719	0.546	0.722	0.719	0.718	0.903	0.803	0.837	0.820	0.796	0.585	0.633	0.608	0.891	0.741	0.571	0.645
Fold 3	0.699	0.512	0.706	0.699	0.697	0.907	0.801	0.825	0.813	0.803	0.551	0.642	0.593	0.867	0.723	0.486	0.581
Fold 4	0.704	0.521	0.714	0.704	0.701	0.902	0.812	0.831	0.821	0.816	0.561	0.679	0.614	0.857	0.721	0.443	0.549
Fold 5	0.711	0.535	0.714	0.711	0.711	0.900	0.815	0.820	0.818	0.808	0.579	0.642	0.609	0.880	0.684	0.557	0.614
Fold 6	0.691	0.498	0.696	0.691	0.689	0.891	0.787	0.820	0.804	0.798	0.545	0.615	0.578	0.870	0.714	0.500	0.588
Fold 7	0.694	0.504	0.698	0.694	0.691	0.904	0.808	0.832	0.820	0.806	0.544	0.624	0.581	0.871	0.673	0.471	0.555
Fold 8	0.702	0.513	0.707	0.702	0.697	0.891	0.779	0.844	0.810	0.795	0.570	0.633	0.600	0.862	0.750	0.471	0.579
Fold 9	0.682	0.484	0.691	0.682	0.679	0.892	0.786	0.814	0.800	0.794	0.535	0.639	0.582	0.859	0.705	0.437	0.539
Fold10	0.705	0.519	0.715	0.705	0.701	0.896	0.782	0.843	0.812	0.805	0.565	0.642	0.601	0.871	0.791	0.478	0.596
Mean	0.701	0.514	0.707	0.701	0.698	0.898	0.798	0.830	0.813	0.802	0.558	0.639	0.596	0.870	0.724	0.489	0.582
RFC																	
Fold 1	0.725	0.551	0.745	0.725	0.721	0.907	0.800	0.843	0.821	0.824	0.583	0.706	0.639	0.896	0.868	0.471	0.611
Fold 2	0.722	0.551	0.722	0.722	0.720	0.910	0.793	0.831	0.812	0.825	0.605	0.633	0.619	0.900	0.737	0.600	0.661
Fold 3	0.730	0.560	0.742	0.730	0.726	0.915	0.792	0.849	0.820	0.828	0.608	0.697	0.650	0.891	0.833	0.500	0.625
Fold 4	0.716	0.536	0.742	0.716	0.708	0.913	0.807	0.831	0.819	0.839	0.577	0.752	0.653	0.887	0.844	0.386	0.529
Fold 5	0.743	0.578	0.760	0.743	0.738	0.911	0.790	0.856	0.822	0.825	0.627	0.725	0.672	0.896	0.897	0.500	0.642
Fold 6	0.714	0.534	0.726	0.714	0.712	0.903	0.778	0.820	0.799	0.827	0.584	0.670	0.624	0.893	0.822	0.529	0.643
Fold 7	0.725	0.548	0.738	0.725	0.719	0.917	0.784	0.868	0.824	0.832	0.602	0.679	0.638	0.889	0.842	0.457	0.593
Fold 8	0.725	0.548	0.741	0.725	0.718	0.901	0.774	0.862	0.816	0.826	0.610	0.688	0.647	0.890	0.865	0.457	0.598
Fold 9	0.714	0.531	0.718	0.714	0.706	0.906	0.778	0.862	0.818	0.814	0.597	0.657	0.626	0.889	0.762	0.451	0.566
Fold10	0.714	0.531	0.723	0.714	0.707	0.909	0.772	0.855	0.811	0.825	0.595	0.661	0.626	0.897	0.805	0.465	0.589
Mean	0.723	0.547	0.736	0.723	0.718	0.909	0.787	0.848	0.816	0.827	0.599	0.687	0.639	0.893	0.828	0.482	0.606
LGBM																	
Fold 1	0.745	0.586	0.746	0.745	0.741	0.929	0.814	0.867	0.840	0.859	0.636	0.688	0.661	0.923	0.760	0.543	0.633
Fold 2	0.751	0.593	0.750	0.751	0.746	0.933	0.799	0.886	0.840	0.854	0.651	0.651	0.651	0.926	0.788	0.586	0.672
Fold 3	0.739	0.576	0.744	0.739	0.735	0.935	0.811	0.855	0.833	0.872	0.629	0.716	0.670	0.923	0.761	0.500	0.603
Fold 4	0.754	0.599	0.759	0.754	0.748	0.938	0.819	0.873	0.845	0.869	0.645	0.734	0.687	0.921	0.795	0.500	0.614
Fold 5	0.746	0.586	0.749	0.746	0.743	0.929	0.804	0.862	0.832	0.863	0.632	0.679	0.655	0.929	0.800	0.571	0.667
Fold 6	0.737	0.570	0.738	0.737	0.733	0.930	0.777	0.856	0.815	0.873	0.655	0.679	0.667	0.923	0.776	0.543	0.639
Fold 7	0.746	0.585	0.747	0.746	0.741	0.935	0.807	0.874	0.839	0.866	0.638	0.679	0.658	0.921	0.776	0.543	0.639
Fold 8	0.749	0.588	0.751	0.749	0.744	0.927	0.790	0.880	0.833	0.869	0.652	0.670	0.661	0.927	0.813	0.557	0.661
Fold 9	0.723	0.546	0.722	0.723	0.717	0.923	0.767	0.868	0.815	0.854	0.623	0.611	0.617	0.916	0.765	0.549	0.639
Fold10	0.746	0.584	0.749	0.746	0.741	0.931	0.780	0.873	0.824	0.863	0.661	0.679	0.670	0.926	0.813	0.549	0.655
Mean	0.743	0.581	0.746	0.743	0.739	0.931	0.797	0.870	0.832	0.864	0.642	0.679	0.659	0.923	0.785	0.544	0.642
Catboost																	
Fold 1	0.713	0.532	0.735	0.713	0.708	0.905	0.803	0.837	0.820	0.819	0.566	0.706	0.629	0.899	0.833	0.429	0.566
Fold 2	0.719	0.544	0.731	0.719	0.718	0.911	0.802	0.831	0.817	0.819	0.575	0.670	0.619	0.902	0.804	0.529	0.638
Fold 3	0.725	0.550	0.738	0.725	0.720	0.914	0.797	0.849	0.822	0.835	0.594	0.697	0.641	0.895	0.825	0.471	0.600
Fold 4	0.707	0.524	0.723	0.707	0.701	0.913	0.812	0.831	0.821	0.834	0.565	0.716	0.632	0.886	0.757	0.400	0.523
Fold 5	0.717	0.540	0.725	0.717	0.713	0.908	0.809	0.838	0.824	0.818	0.581	0.688	0.630	0.894	0.750	0.471	0.578
Fold 6	0.699	0.510	0.712	0.699	0.696	0.899	0.784	0.826	0.805	0.821	0.558	0.661	0.605	0.895	0.780	0.457	0.577
Fold 7	0.702	0.513	0.717	0.702	0.696	0.915	0.797	0.844	0.820	0.828	0.553	0.670	0.606	0.890	0.784	0.414	0.542
Fold 8	0.697	0.501	0.707	0.697	0.689	0.899	0.762	0.844	0.801	0.821	0.569	0.642	0.603	0.893	0.789	0.429	0.556
Fold 9	0.702	0.513	0.716	0.702	0.696	0.899	0.783	0.844	0.813	0.810	0.563	0.667	0.610	0.889	0.789	0.423	0.550
Fold10	0.702	0.513	0.715	0.702	0.696	0.906	0.778	0.843	0.809	0.818	0.567	0.661	0.610	0.895	0.795	0.437	0.564
Mean	0.708	0.524	0.722	0.708	0.703	0.907	0.793	0.839	0.815	0.822	0.569	0.678	0.619	0.894	0.791	0.446	0.569
ANN																	
Fold 1	0.710	0.534	0.716	0.710	0.710	0.915	0.819	0.819	0.819	0.816	0.568	0.651	0.607	0.886	0.704	0.543	0.613
Fold 2	0.701	0.521	0.703	0.701	0.701	0.907	0.799	0.813	0.806	0.805	0.565	0.596	0.580	0.892	0.689	0.600	0.641
Fold 3	0.699	0.516	0.703	0.699	0.699	0.918	0.819	0.819	0.819	0.807	0.553	0.624	0.586	0.885	0.661	0.529	

Table S4 The best hyperparameters for the six multi-classification ML models

Multi-classification models	Hyperparameters	Optimal value
LR	C	0.6
	class_weight	None
	penalty	l2
SVM	Multi-class	auto
	C	0.8
	kernel	rbf
	probability	True
	gamma	auto
RFC	n_estimators	450
	max_depth	6
	min_samples_leaf	12
	Min_samples_split	8
	num_class	3
LGBM	objective	multiclass
	learning_rate	0.003
	num_leaves	12
	metric	multi-logloss
	n_estimators	450
Catboost	iterations	400
	learning_rate	0.003
	depth	6
	loss_function	MultiClass
ANN	Dense	(128,); (64,); (3,)
	activation	'relu'; 'relu'; 'softmax'
	Dropout	(0.2); (0.2)
	EarlyStopping	Monitor = 'val_loss'; min_delta = 0; patience = 50; verbose = 1

Notes: ML, machine learning; LR, logistic regression; SVM, support vector machine; LGBM, light gradient boosting machine; Catboost, categorical boosting; ANN, artificial neural network.

2 Supplemental Figures

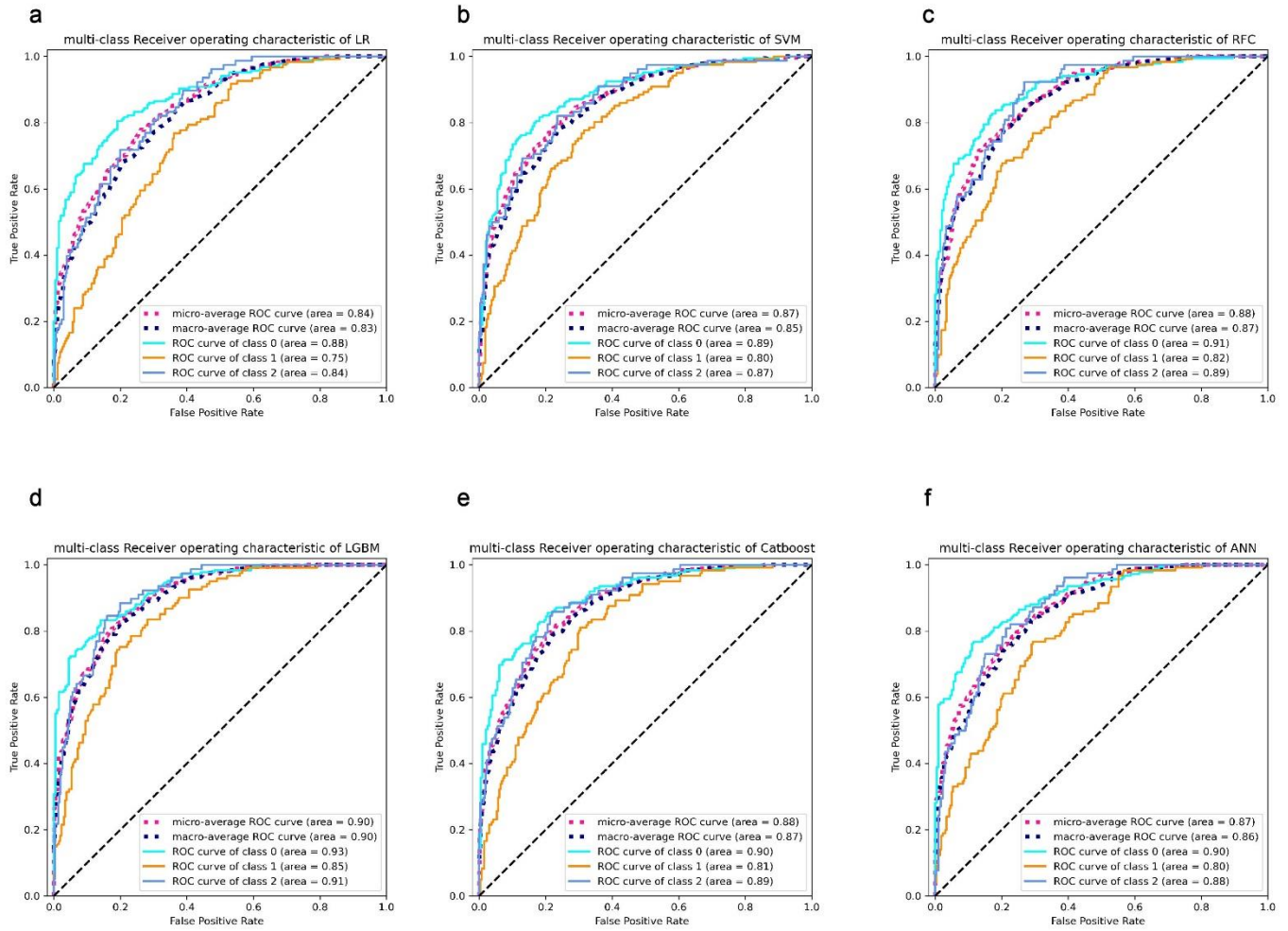


Figure S1 ROC curve of the six multi-classification ML models in the training cohort. ROC, receiver operating characteristic; ML, machine learning.

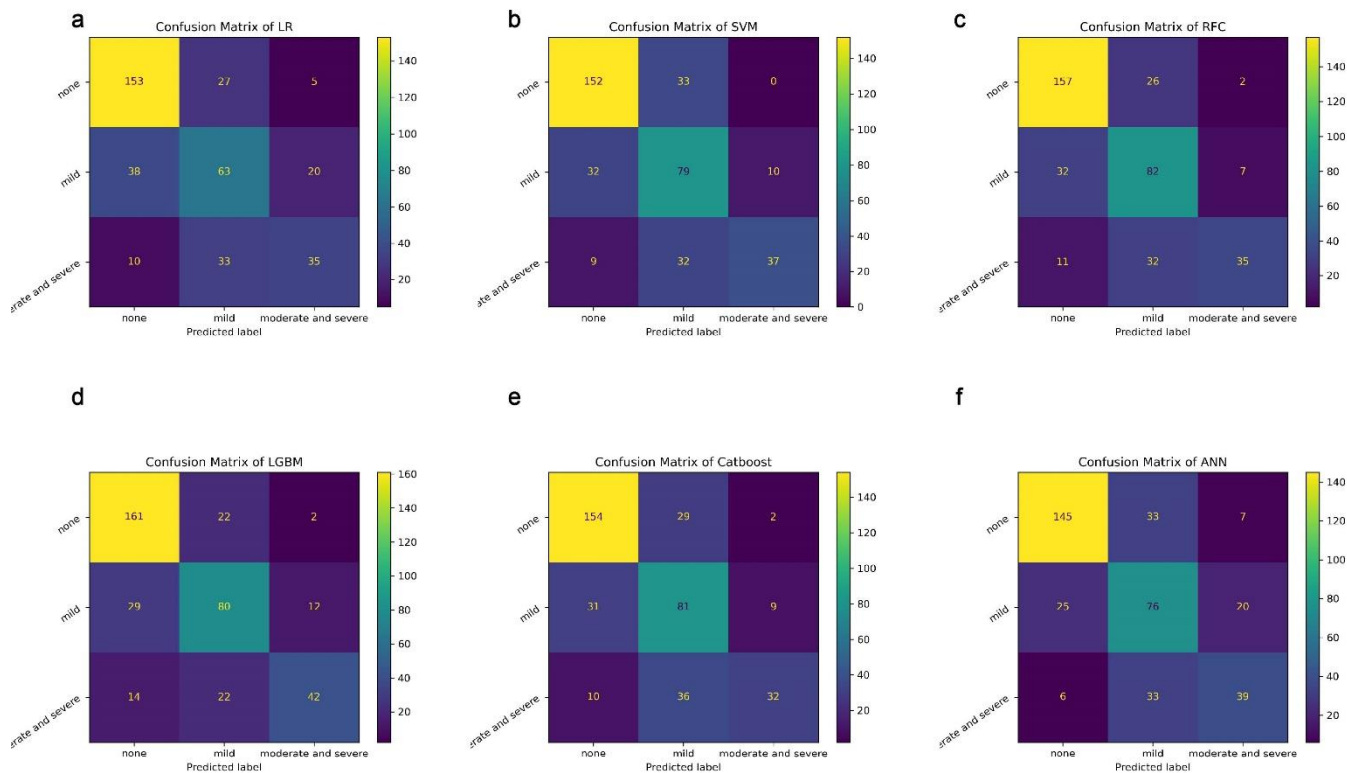


Figure S2 Confusion matrix of the six multi-classification ML models in the training cohort. ML, machine learning model; LR, logistic regression; SVM, support vector machine; RFC, random forest classifier; LGBM, light gradient boosting machine; Catboost, categorical boosting; ANN, artificial neural network.

Real-Time Depression Prediction

Fatigue

Less than seven days within two weeks

Sleep time(h/day)

1 5 12

1 3 5 7 9 11 12

Platelet($10^9/L$)

4 100 715

4 76 148 220 292 364 436 508 580 652 715

Lymphocyte($10^9/L$)

0.14 1 3.38

0.14 0.49 0.84 1.19 1.54 1.89 2.24 2.59 2.94 3.2688

Anti-Ro52

Positive

Predict Clear

The probability of Depression:

[1] 0.563

The risk group:

[1] "Mild depression"

Figure S3 The screenshot of the web-based tool. As shown in the figure, the basic information of a patient is that he or she experienced fatigue for less than seven days within two weeks, slept for 5 hours/day, had a platelet count of 100 ($10^9/L$), lymphocyte count of 1 ($10^9/L$), and a positive

anti-RO52 antibody. Our ANN model predicts that this patient had a probability value of 0.563 for the occurrence of mild depression.