Supplementary Information

C Model comparison

After preprocessing the data, we compared the performance of 4 ensembles based on 4 types of base learners on the validation sets: Logistic Regression (LR), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Light Gradient Boosting Model (LGBM). The models were compared using the AUROC and AUPRC and the results are shown in Table B. We selected the ensemble that achieved the highest AUROC on the validation set.

Table B: Performance comparison for the different ensembles on the validation sets. Best performance is shown in bold.

Complication	Models	Validation Set A		Validation Set B	
		AUROC	**AUPRC**	**AUROC**	**AUPRC**
SBI	LR	**0.895**	0.411	**0.980**	0.596
	KNN	0.813	0.285	0.861	0.471
	SVM	0.853	0.321	0.965	0.543
	LGBM	0.888	0.382	0.934	0.395
AKI	LR	**0.888**	0.411	**0.892**	0.272
	KNN	0.777	0.234	0.770	0.166
	SVM	0.814	0.138	0.830	0.221
	LGBM	0.854	0.404	0.873	0.166
ARDS	LR	**0.917**	0.379	0.942	0.396
	KNN	0.814	0.245	0.809	0.290
	SVM	0.868	0.217	0.921	0.412
	LGBM	0.912	0.352	**0.949**	0.417