Supplementary table legends:

Supplementary Table S1. Comparison of independent testing results between IL10-Stack and other existing methods.

	Sampling method	Machine learning algorithm	Cross validation						Independent test					
			ACC	MCC	Sn	Sp	AUC	P	ACC	MCC	Sn	Sp	AUC	P
Under-sampling	RUS [b]	SVM	0.879	0.758	0.842	0.913	0.944	0.901	0.919	0.838	0.907	**0.931** [a]	0.966	0.933
	RUS	LGBM	0.872	0.744	0.865	0.879	0.933	0.870	0.904	0.809	0.921	0.886	0.950	0.896
	RUS	Stack	0.881	0.765	0.895	0.868	0.913	0.867	0.911	0.823	0.935	0.886	0.938	0.898
	CC [b]	SVM	0.711	0.421	0.692	0.728	0.797	0.709	0.685	0.392	0.581	0.804	0.788	0.772
	CC	LGBM	0.785	0.570	0.758	0.811	0.842	0.793	0.761	0.523	0.752	0.772	0.840	0.790
	CC	Stack	0.785	0.575	0.768	0.801	0.788	0.790	0.751	0.503	0.743	0.761	0.755	0.780
	NM [b]	SVM	0.788	0.577	0.772	0.805	0.845	0.791	0.751	0.503	0.743	0.761	0.808	0.780
	NM	LGBM	0.765	0.530	0.723	0.805	0.857	0.780	0.726	0.457	0.686	0.772	0.795	0.774
	NM	Stack	0.788	0.579	0.737	0.838	0.789	0.812	0.726	0.457	0.686	0.772	0.725	0.774
Over-sampling	ROS [b]	SVM	0.867	0.737	0.821	0.912	0.938	0.900	0.903	0.806	0.902	0.905	**0.968**	0.914
	ROS	LGBM	0.869	0.739	0.851	0.887	0.940	0.879	0.920	0.839	0.933	0.905	0.966	0.917
	ROS	Stack	0.877	0.756	0.856	0.898	0.908	0.891	0.920	0.839	0.937	0.900	0.956	0.913
	ADASYN [b]	SVM	0.888	0.776	0.866	0.909	0.949	0.899	**0.921**	**0.842**	0.912	**0.931**	0.964	**0.934**
	ADASYN	LGBM	0.860	0.720	0.845	0.875	0.924	0.863	0.890	0.780	0.903	0.876	0.938	0.886
	ADASYN	Stack	0.882	0.764	0.873	0.890	0.906	0.883	0.916	0.833	**0.940**	0.891	0.939	0.902
	SMOTE [b]	SVM	0.881	0.762	0.845	**0.915**	**0.951**	**0.905**	0.892	0.784	0.871	0.915	0.952	0.920
	SMOTE	LGBM	0.866	0.732	0.837	0.894	0.922	0.883	0.884	0.769	0.875	0.895	0.948	0.903
	SMOTE	Stack	**0.897**	**0.796**	**0.918**	0.877	0.913	0.878	0.910	0.820	0.933	0.885	0.920	0.901

[a] The best performance values are indicated in bold and underlined.

[b] RUS: random under-sampling; CC: cluster centroids; NM: near miss; ROS: random over-sampling; ADASYN: adaptive synthetic; SMOTE: synthetic minority over-sampling technique.

We tried six different methods, including SMOTE, to balance the data for various models. Our approaches mainly fall into two categories: oversampling and under-sampling. Oversampling involves increasing the instances of the minority class, while under-sampling involves decreasing the instances of the majority class. Below, we provide these different methods' names and principles.

i) Under-sampling

- Random Under-Sampling (RUS)

Principle: Randomly select and remove majority class samples.

- Cluster Centroids (CC)

Principle: By clustering the majority class samples and then selecting the cluster centers as new minority class samples, data balance is achieved.

- Near Miss (NM)

Principle: Attempt to retain the majority class samples that are closest to the minority class samples to better capture the characteristics of the minority class.

ii) Over-sampling

- Random Over-Sampling (ROS)

Principle: Randomly duplicate the minority class samples to generate new samples, thus increasing their quantity to match that of the majority class.

- Adaptive Synthetic (ADASYN)

Principle: The number of generated samples is determined based on the ratio of majority class samples in the neighboring samples, reducing data imbalance.

- Synthetic Minority Oversampling Technique (SMOTE)

Principle: Generate minority samples through random linear interpolation between minority sample points and their neighboring points.