

Explainable Classification of Astronomical Uncertain Time Series (Supplementary material)

1 SAST-based methods results

Table 1: SAST<X> models results on the PLAsTiCC dataset average over 3 runs. The starred row is the run that achieved the lowest cross entropy loss.

	Classifier	Precision	Recall	F1 score	LogLoss	Time (h)
SAST	RF	0.59 ± 0.04	0.61 ± 0.00	0.56 ± 0.01	1.29 ± 0.05	16.41 ± 0.52
	RidgeCV	0.56 ± 0.00	0.57 ± 0.00	0.55 ± 0.00	2.22 ± 0.01	16.41 ± 0.52
	XGBoost	0.65 ± 0.01	0.67 ± 0.00	0.63 ± 0.00	1.16 ± 0.01	16.41 ± 0.52
	XGBoost*	0.65	0.67	0.64	1.15	16.23
SASTd	RF	0.58 ± 0.05	0.60 ± 0.00	0.54 ± 0.00	1.33 ± 0.08	12.79 ± 0.84
	RidgeCV	0.56 ± 0.01	0.58 ± 0.01	0.55 ± 0.01	2.20 ± 0.00	12.79 ± 0.84
	XGBoost	0.66 ± 0.02	0.68 ± 0.00	0.64 ± 0.00	1.14 ± 0.00	12.79 ± 0.841
	XGBoost*	0.66	0.68	0.64	1.13	11.84
SASTdc	RF	0.58 ± 0.03	0.61 ± 0.00	0.55 ± 0.00	1.33 ± 0.04	12.99 ± 0.30
	RidgeCV	0.52 ± 0.02	0.54 ± 0.01	0.52 ± 0.01	2.25 ± 0.00	12.99 ± 0.30
	XGBoost	0.66 ± 0.01	0.68 ± 0.00	0.64 ± 0.01	1.14 ± 0.01	12.99 ± 0.30
	XGBoost*	0.65	0.68	0.64	1.14	13.04

2 uSAST-based methods results

Table 2: uSAST<X> models results on the PLAsTiCC dataset average over 3 runs. The row with the star is the run that achieved the lowest cross entropy loss. Subsequence length: from 16 to 32 with a step of 1.

	Classifier	Precision	Recall	F1 score	LogLoss	Time (h)
uSAST	RF	0.62 ± 0.01	0.66 ± 0.01	0.60 ± 0.01	1.20 ± 0.03	53.61 ± 0.10
	RidgeCV	0.33 ± 0.02	0.34 ± 0.02	0.33 ± 0.02	2.98 ± 0.11	53.61 ± 0.10
	XGBoost	0.69 ± 0.00	0.70 ± 0.00	0.66 ± 0.00	1.04 ± 0.01	53.61 ± 0.10
	XGBoost*	0.69	0.70	0.66	1.04	53.56
uSASTd	RF	0.61 ± 0.03	0.64 ± 0.00	0.58 ± 0.00	1.20 ± 0.04	41.58 ± 0.47
	RidgeCV	0.35 ± 0.01	0.37 ± 0.01	0.35 ± 0.01	2.90 ± 0.06	41.58 ± 0.47
	XGBoost	0.70 ± 0.01	0.70 ± 0.00	0.67 ± 0.00	1.06 ± 0.02	41.58 ± 0.47
	XGBoost*	0.71	0.70	0.67	1.04	41.78
uSASTdc	RF	0.60 ± 0.00	0.64 ± 0.01	0.58 ± 0.00	1.21 ± 0.00	40.45 ± 1.00
	RidgeCV	0.37 ± 0.04	0.38 ± 0.03	0.37 ± 0.03	2.78 ± 0.07	40.45 ± 1.00
	XGBoost	0.69 ± 0.01	0.70 ± 0.00	0.66 ± 0.00	1.04 ± 0.02	40.45 ± 1.00
	XGBoost*	0.68	0.70	0.66	1.03	41.16

Table 3: uSAST<X> models results on the PLAsTiCC dataset average over 3 runs. The row with the star is the run that achieved the lowest cross entropy loss. Subsequence length: from 20 to 60 with a step of 10.

	Classifier	Precision	Recall	F1 score	LogLoss	Time (h)
uSAST	RF	0.64 ± 0.01	0.68 ± 0.01	0.62 ± 0.01	1.09 ± 0.01	51.03 ± 0.12
	Ridge	0.39 ± 0.01	0.40 ± 0.01	0.39 ± 0.01	2.82 ± 0.08	51.03 ± 0.12
	XGBoost	0.72 ± 0.01	0.72 ± 0.00	0.69 ± 0.01	0.96 ± 0.01	51.03 ± 0.12
	XGBoost*	0.71	0.72	0.69	0.95	51.17
uSASTd	RF	0.65 ± 0.02	0.67 ± 0.01	0.62 ± 0.01	1.12 ± 0.04	43.49 ± 0.27
	Ridge	0.39 ± 0.02	0.41 ± 0.02	0.40 ± 0.02	2.77 ± 0.07	43.49 ± 0.27
	XGBoost	0.72 ± 0.00	0.73 ± 0.00	0.70 ± 0.01	0.97 ± 0.01	43.49 ± 0.27
	XGBoost*	0.72	0.73	0.69	0.95	43.18
uSASTdc	RF	0.67 ± 0.03	0.67 ± 0.00	0.62 ± 0.00	1.11 ± 0.00	43.52 ± 0.72
	Ridge	0.39 ± 0.02	0.40 ± 0.02	0.39 ± 0.02	2.79 ± 0.02	43.52 ± 0.72
	XGBoost	0.71 ± 0.01	0.72 ± 0.01	0.69 ± 0.01	0.96 ± 0.01	43.52 ± 0.72
	XGBoost*	0.70	0.72	0.68	0.95	44.26

3 Results of state-of-the-art models

Table 4: Performance of state-of-the-art models average over 3 runs. MUSE uses a linear classifier. The row with the star is the run that achieved the lowest cross entropy loss

	Classifier	Precision	Recall	F1 score	LogLoss	Time (h)
MUSE	-	0.71 ± 0.01	0.73 ± 0.01	0.71 ± 0.01	1.78 ± 0.03	3.36 ± 0.04
ROCKET	RF	0.75 ± 0.01	0.75 ± 0.00	0.72 ± 0.00	0.94 ± 0.03	0.05 ± 0.00
	Ridge	0.71 ± 0.01	0.71 ± 0.01	0.70 ± 0.01	2.07 ± 0.00	0.05 ± 0.00
	XGBoost	0.77 ± 0.00	0.77 ± 0.00	0.75 ± 0.00	0.82 ± 0.01	0.05 ± 00
	XGBoost*	0.78	0.77	0.75	0.81	0.05
XEM	-	0.69 ± 0.01	0.71 ± 0.00	0.69 ± 0.00	-	12.24 ± 0.46