

Full Study 1 Individual Dataset Results

Table 1. Correct detection results

Features	Classifier	SMOTE	Kappa	Correct			Incorrect		
				F1	Prec	Recall	F1	Prec	Recall
Chance Baseline		N	0.00	0.90	0.91	0.90	0.11	0.12	0.11
Response Time Baseline	Gradient Boosted	Y	0.25	0.93	0.90	0.98	0.30	0.56	0.21
Global	Gradient Boosted	Y	0.26	0.91	0.91	0.92	0.36	0.33	0.49
Local	XGBoost	Y	0.52	0.96	0.93	0.99	0.58	0.63	0.60
Global + Local	XGBoost	Y	0.55	0.96	0.95	0.96	0.60	0.68	0.54

Table 2. TUT detection Results

Feature Set	Classifier	SMOTE	Kappa	Weighted F1 Score	F1_TUT	Prec. TUT	Recall TUT
Chance Baseline		N	0.00	0.53	0.25	0.25	0.25
Response Time Baseline	Logistic Regression	Y	0.12	0.66	0.27	0.38	0.21
Global	Gradient Boosted	Y	0.13	0.67	0.33	0.37	0.32
Local	Gradient Boosted	Y	0.17	0.69	0.33	0.40	0.30
Global + Local	Gradient Boosted	Y	0.19	0.70	0.37	0.33	0.42

Table 3. Convergent validity

	TUT Ground Truth	Correctness Ground Truth	TUT Predicted
Correctness Ground Truth	-0.207*		
TUT Predicted	0.157	-0.242*	
Correctness Predicted	-0.243*	0.662*	-0.267*

Full Study 2 Individual Dataset Results

Table 4. Correct detection results

Features	Classifier	SMOTE	Kappa	Correct			Incorrect		
				F1	Prec	Recall	F1	Prec	Recall
Chance Baseline		N	0	0.90	0.90	0.90	0.10	0.10	0.10
Response Time Baseline	Gradient Boosted	Y	0.08	0.90	0.91	0.89	0.09	0.13	0.09
Global	Gradient Boosted	Y	0.06	0.91	0.90	0.94	0.11	0.15	0.08
Local	XGBoost	Y	0.58	0.97	0.94	0.99	0.61	0.97	0.44
Global + Local	XGBoost	Y	0.50	0.96	0.93	0.99	0.53	0.88	0.38

Table 4. TUT detection results

Feature Set	Classifier	SMOTE	Kappa	Weighted F1 Score	F1_TUT	Prec. TUT	Recall TUT
Chance Baseline		N	0	0.54	0.23	0.23	0.23
Response Time Baseline	Logistic Regression	Y	0.10	0.68	0.29	0.31	0.27
Global	Logistic Regression	Y	0.07	0.67	0.31	0.27	0.39
Local	XG Boost	Y	0.08	0.69	0.29	0.26	0.34
Global + Local	XG Boost	Y	0.12	0.70	0.33	0.24	0.53

Table 6. Convergent validity

	TUT Ground Truth	Correctness Ground Truth	TUT Predicted
Correctness Ground Truth	-0.05		
TUT Predicted	0.210*	-0.133	
Correctness Predicted	0.01	0.802*	-0.150 *