

Input Type	Model	Metrics	Single-task			Multi-task		
			Artifact	Valid pulse	Macro average	Artifact	Valid pulse	Macro average
Multi-pulse	CNN-GRU	Precision	0.9787	0.9106	0.9447	0.9738	0.8901	0.9319
		Recall	0.9737	0.9267	0.9502	0.9676	0.9098	0.9387
		F1-score	0.9762	0.9186	0.9474	0.9707	0.8998	0.9352
	GRU	Precision	0.9805	0.8752	0.9278	0.9706	0.8824	0.9265
		Recall	0.9615	0.9337	0.9476	0.9654	0.8987	0.9321
		F1-score	0.9709	0.9035	0.9372	0.9680	0.8905	0.9292
	ResNet	Precision	0.9779	0.8701	0.9240	0.9658	0.8564	0.9111
		Recall	0.9601	0.9249	0.9425	0.9572	0.8828	0.9200
		F1-score	0.9689	0.8967	0.9328	0.9615	0.8694	0.9154
Single-pulse	CNN-GRU	Precision	0.9831	0.8746	0.9289	0.9731	0.8706	0.9219
		Recall	0.9609	0.9429	0.9519	0.9610	0.9082	0.9346
		F1-score	0.9719	0.9075	0.9397	0.9670	0.8890	0.9280
	GRU	Precision	0.9879	0.8762	0.9320	0.9682	0.8485	0.9084
		Recall	0.9609	0.9591	0.9600	0.9540	0.8915	0.9228
		F1-score	0.9742	0.9158	0.9450	0.9611	0.8695	0.9153
	ResNet	Precision	0.9823	0.8772	0.9298	0.9741	0.8738	0.9239
		Recall	0.9620	0.9401	0.9510	0.9620	0.9113	0.9366
		F1-score	0.9720	0.9076	0.9398	0.9680	0.8922	0.9301

Table S1. Artifact detection performances for different input types, models, and designs. The best performance for each task using the same design is bolded.

Input Type	Model	Metrics	Single-task				Multi-task			
			Compliant	Multi-peak noncompliant	Single-peak noncompliant	Macro average	Compliant	Multi-peak noncompliant	Single-peak noncompliant	Macro average
Multi-pulse	CNN-GRU	Precision	0.9624	0.889	0.8374	0.8963	0.9167	0.9198	0.8175	0.8846
		Recall	0.9331	0.8387	0.9599	0.9106	0.9451	0.7623	0.9611	0.8895
		F1-score	0.9475	0.8631	0.8945	0.9017	0.9307	0.8336	0.8835	0.8826
	GRU	Precision	0.9195	0.9173	0.8442	0.8937	0.9478	0.8757	0.7486	0.8574
		Recall	0.9500	0.7930	0.9458	0.8962	0.8978	0.7672	0.9658	0.8769
		F1-score	0.9345	0.8506	0.8921	0.8924	0.9221	0.8179	0.8435	0.8611
	ResNet	Precision	0.9484	0.8804	0.8189	0.8826	0.8587	0.8147	0.7883	0.8206
		Recall	0.9299	0.8080	0.9493	0.8957	0.8858	0.6650	0.9399	0.8302
		F1-score	0.9390	0.8427	0.8793	0.8870	0.8721	0.7323	0.8575	0.8206
Single-pulse	CNN-GRU	Precision	0.9407	0.8574	0.7670	0.8550	0.9327	0.7911	0.8236	0.8491
		Recall	0.8717	0.8146	0.9434	0.8766	0.8293	0.8529	0.9304	0.8708
		F1-score	0.9049	0.8355	0.8461	0.8622	0.8780	0.8208	0.8738	0.8575
	GRU	Precision	0.9421	0.9307	0.7387	0.8705	0.9407	0.8889	0.7394	0.8563
		Recall	0.8940	0.8038	0.9634	0.8871	0.8706	0.8113	0.9505	0.8775
		F1-score	0.9174	0.8626	0.8362	0.8721	0.9043	0.8483	0.8318	0.8615
	ResNet	Precision	0.9492	0.8248	0.7479	0.8406	0.9056	0.8748	0.7048	0.8284
		Recall	0.8222	0.8296	0.9587	0.8702	0.8456	0.7722	0.9233	0.8471
		F1-score	0.8811	0.8272	0.8403	0.8495	0.8746	0.8203	0.7994	0.8314

Table S2. Morphology classification performances for different input types, models, and designs. The best performance for each task using the same design is bolded.