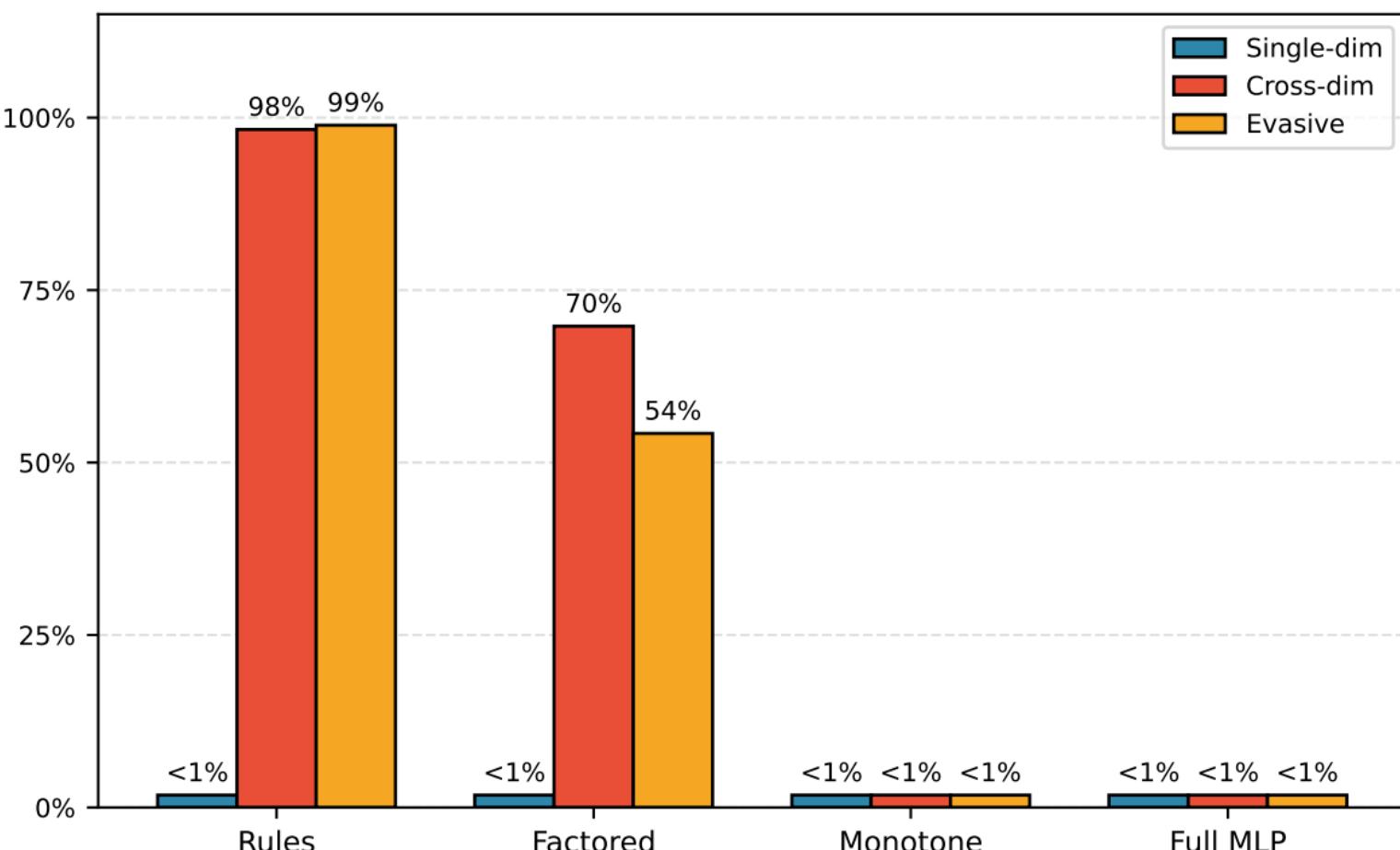
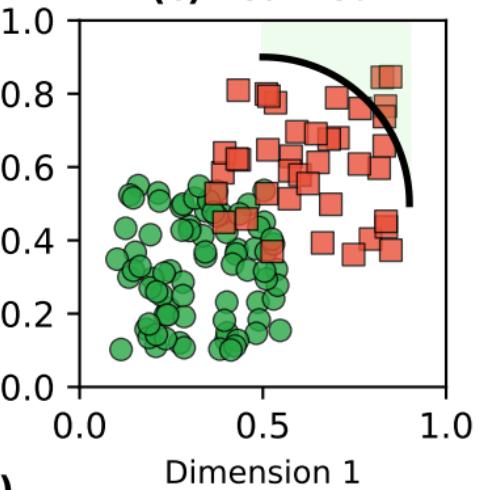
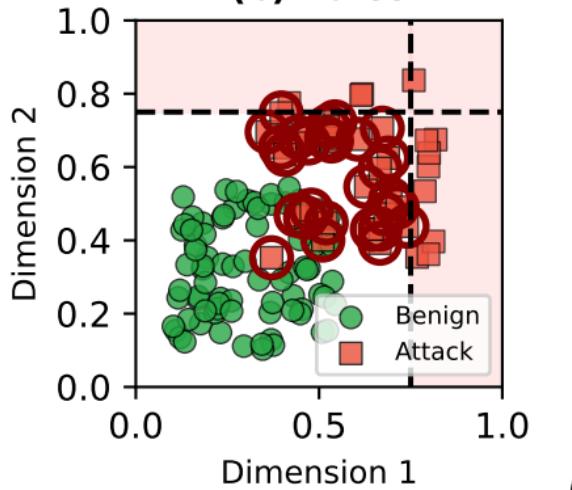


(a) FAR by Attack Type (1% FRR)**(c) Learned****(b) Rules****(d) Summary (1% FRR)**

Model	FAR	FRR	Acc	ECE
Rules	0.67	0.01	0.73	0.17
Factored	0.42	0.01	0.83	0.23
Monotone	0.00	0.01	0.99	0.01
Full MLP	0.00	0.01	0.99	0.01