Estimating the Limits of Organism-Specific Training for Epitope Prediction Results Tables

J. Ashford, A. Ekárt, F. Campelo

September 25, 2023

1 Full results tables

Tables 1–3 document the numerical values of performance estimates for the models tested in this work, together with sample standard errors.

Table 1: Point estimates and standard errors of mean performance for models trained on the *Onchocerca volvulus* data. Row *Hybrid-B*: 0 Peptides corresponds to the "heterogeneous" case, with 0 organism- specific and 1,000 non-target pathogen peptides in the training set.

Training	Peptides	AUC	Bal. Accuracy	MCC	PPV	NPV	Sensitivity
OrgSpec	20	0.70 ± 0.000	0.65 ± 0.000	0.30 ± 0.000	0.67 ± 0.000	0.63 ± 0.000	0.65 ± 0.000
	40	0.74 ± 0.000	0.67 ± 0.000	0.35 ± 0.001	0.69 ± 0.000	0.66 ± 0.000	0.67 ± 0.000
	60	0.76 ± 0.000	0.69 ± 0.000	0.38 ± 0.001	0.71 ± 0.000	0.67 ± 0.000	0.68 ± 0.001
	80	0.77 ± 0.000	0.70 ± 0.000	0.40 ± 0.001	0.70 ± 0.000	0.69 ± 0.001	0.72 ± 0.001
	100	0.77 ± 0.001	0.70 ± 0.001	0.41 ± 0.001	0.71 ± 0.001	0.69 ± 0.001	0.72 ± 0.001
	150	0.78 ± 0.001	0.71 ± 0.001	0.42 ± 0.001	0.72 ± 0.001	0.70 ± 0.001	0.72 ± 0.001
	200	0.79 ± 0.001	0.72 ± 0.001	0.44 ± 0.001	0.73 ± 0.001	0.71 ± 0.001	0.74 ± 0.001
	250	0.80 ± 0.001	0.72 ± 0.001	0.44 ± 0.001	0.73 ± 0.001	0.72 ± 0.001	0.74 ± 0.001
	300	0.80 ± 0.001	0.73 ± 0.001	0.45 ± 0.001	0.73 ± 0.001	0.72 ± 0.001	0.75 ± 0.001
	400	0.81 ± 0.001	0.73 ± 0.001	0.47 ± 0.002	0.74 ± 0.001	0.73 ± 0.001	0.76 ± 0.001
	500	0.81 ± 0.001	0.74 ± 0.001	0.47 ± 0.003	0.74 ± 0.002	0.74 ± 0.002	0.76 ± 0.002
Hybrid-A	20	0.65 ± 0.000	0.60 ± 0.000	0.21 ± 0.000	0.62 ± 0.000	0.61 ± 0.000	0.63 ± 0.001
	40	0.68 ± 0.000	0.62 ± 0.000	0.25 ± 0.001	0.64 ± 0.000	0.61 ± 0.000	0.62 ± 0.001
	60	0.69 ± 0.000	0.63 ± 0.000	0.27 ± 0.001	0.65 ± 0.000	0.63 ± 0.001	0.64 ± 0.001
	80	0.70 ± 0.001	0.65 ± 0.001	0.30 ± 0.001	0.65 ± 0.001	0.65 ± 0.001	0.68 ± 0.002
	100	0.72 ± 0.001	0.65 ± 0.001	0.32 ± 0.001	0.67 ± 0.001	0.66 ± 0.001	0.68 ± 0.002
	150	0.73 ± 0.001	0.66 ± 0.001	0.33 ± 0.002	0.67 ± 0.001	0.66 ± 0.001	0.68 ± 0.003
	200	0.74 ± 0.001	0.68 ± 0.001	0.35 ± 0.002	0.68 ± 0.001	0.67 ± 0.002	0.70 ± 0.003
	250	0.76 ± 0.001	0.69 ± 0.001	0.38 ± 0.002	0.69 ± 0.001	0.70 ± 0.001	0.74 ± 0.003
	300	0.76 ± 0.001	0.69 ± 0.001	0.39 ± 0.002	0.69 ± 0.002	0.71 ± 0.002	0.75 ± 0.004
	400	0.77 ± 0.001	0.70 ± 0.001	0.40 ± 0.002	0.70 ± 0.001	0.71 ± 0.002	0.75 ± 0.004
	500	0.77 ± 0.001	0.71 ± 0.002	0.42 ± 0.003	0.70 ± 0.003	0.72 ± 0.003	0.75 ± 0.006
Hybrid-B	0	0.55 ± 0.001	0.53 ± 0.001	0.07 ± 0.002	0.54 ± 0.001	0.52 ± 0.001	0.63 ± 0.003
	20	0.56 ± 0.000	0.54 ± 0.000	0.09 ± 0.000	0.55 ± 0.000	0.54 ± 0.000	0.64 ± 0.000
	40	0.58 ± 0.000	0.56 ± 0.000	0.12 ± 0.000	0.57 ± 0.000	0.55 ± 0.000	0.64 ± 0.001
	60	0.60 ± 0.000	0.57 ± 0.000	0.14 ± 0.001	0.57 ± 0.000	0.57 ± 0.000	0.67 ± 0.001
	80	0.60 ± 0.001	0.57 ± 0.000	0.15 ± 0.001	0.58 ± 0.000	0.57 ± 0.001	0.67 ± 0.001
	100	0.62 ± 0.001	0.58 ± 0.001	0.17 ± 0.001	0.59 ± 0.001	0.59 ± 0.001	0.68 ± 0.002
	150	0.65 ± 0.001	0.61 ± 0.001	0.22 ± 0.001	0.61 ± 0.001	0.61 ± 0.001	0.68 ± 0.002
	200	0.68 ± 0.001	0.63 ± 0.001	0.26 ± 0.002	0.63 ± 0.001	0.64 ± 0.001	0.71 ± 0.003
	250	0.69 ± 0.002	0.64 ± 0.001	0.28 ± 0.002	0.64 ± 0.001	0.64 ± 0.002	0.70 ± 0.003
	300	0.72 ± 0.002	0.65 ± 0.002	0.31 ± 0.003	0.65 ± 0.001	0.67 ± 0.003	0.73 ± 0.004
	400	0.76 ± 0.001	0.69 ± 0.001	0.38 ± 0.002	0.69 ± 0.002	0.70 ± 0.002	0.74 ± 0.004
	500	0.77 ± 0.002	0.70 ± 0.002	0.40 ± 0.003	0.69 ± 0.002	0.72 ± 0.003	0.77 ± 0.005

Table 2: Point estimates and standard errors of mean performance for models trained on the Epstein-Barr Virus data. Row *Hybrid-B*: 0 Peptides corresponds to the "heterogeneous" case, with 0 organism- specific and 1,000 non-target pathogen peptides in the training set.

Training	Peptides	AUC	Bal. Accuracy	MCC	PPV	NPV	Sensitivity
OrgSpec	20	0.57 ± 0.001	0.53 ± 0.000	0.11 ± 0.001	0.68 ± 0.000	0.52 ± 0.001	0.94 ± 0.000
	40	0.59 ± 0.001	0.55 ± 0.001	0.13 ± 0.001	0.69 ± 0.000	0.52 ± 0.002	0.92 ± 0.001
	60	0.61 ± 0.001	0.55 ± 0.001	0.14 ± 0.001	0.69 ± 0.000	0.53 ± 0.002	0.93 ± 0.001
	80	0.61 ± 0.002	0.55 ± 0.001	0.16 ± 0.002	0.69 ± 0.001	0.56 ± 0.003	0.93 ± 0.001
	100	0.61 ± 0.002	0.56 ± 0.001	0.17 ± 0.002	0.69 ± 0.001	0.55 ± 0.003	0.92 ± 0.001
	150	0.63 ± 0.003	0.56 ± 0.002	0.19 ± 0.004	0.70 ± 0.001	0.60 ± 0.005	0.93 ± 0.002
	200	0.64 ± 0.003	0.57 ± 0.002	0.21 ± 0.004	0.70 ± 0.001	0.62 ± 0.005	0.94 ± 0.002
	250	0.65 ± 0.005	0.57 ± 0.003	0.21 ± 0.007	0.70 ± 0.002	0.62 ± 0.006	0.94 ± 0.003
	300	0.66 ± 0.005	0.58 ± 0.003	0.23 ± 0.007	0.71 ± 0.002	0.63 ± 0.008	0.93 ± 0.002
	400	0.67 ± 0.005	0.59 ± 0.003	0.26 ± 0.008	0.71 ± 0.002	0.66 ± 0.011	0.93 ± 0.004
	500	0.68 ± 0.011	0.58 ± 0.005	0.23 ± 0.011	0.71 ± 0.003	0.62 ± 0.007	0.93 ± 0.003
Hybrid-A	20	0.51 ± 0.001	0.52 ± 0.000	0.04 ± 0.001	0.67 ± 0.000	0.38 ± 0.001	0.79 ± 0.001
	40	0.51 ± 0.001	0.52 ± 0.001	0.04 ± 0.001	0.67 ± 0.000	0.37 ± 0.001	0.77 ± 0.002
	60	0.52 ± 0.001	0.53 ± 0.001	0.06 ± 0.002	0.68 ± 0.001	0.39 ± 0.002	0.80 ± 0.002
	80	0.53 ± 0.002	0.52 ± 0.001	0.06 ± 0.002	0.68 ± 0.001	0.38 ± 0.002	0.79 ± 0.002
	100	0.53 ± 0.002	0.53 ± 0.001	0.08 ± 0.003	0.69 ± 0.001	0.41 ± 0.002	0.82 ± 0.002
	150	0.54 ± 0.004	0.54 ± 0.002	0.10 ± 0.004	0.69 ± 0.001	0.43 ± 0.004	0.82 ± 0.003
	200	0.52 ± 0.005	0.53 ± 0.003	0.07 ± 0.006	0.68 ± 0.001	0.40 ± 0.006	0.81 ± 0.004
	250	0.54 ± 0.005	0.54 ± 0.003	0.09 ± 0.006	0.69 ± 0.001	0.42 ± 0.005	0.81 ± 0.003
	300	0.56 ± 0.005	0.55 ± 0.004	0.10 ± 0.009	0.69 ± 0.002	0.42 ± 0.008	0.81 ± 0.003
	400	0.56 ± 0.007	0.56 ± 0.004	0.14 ± 0.009	0.70 ± 0.002	0.46 ± 0.009	0.82 ± 0.008
	500	0.56 ± 0.014	0.55 ± 0.006	0.11 ± 0.014	0.69 ± 0.004	0.43 ± 0.012	0.81 ± 0.004
Hybrid-B	0	0.37 ± 0.001	0.41 ± 0.001	-0.16 ± 0.001	0.59 ± 0.001	0.26 ± 0.001	0.46 ± 0.002
	20	0.38 ± 0.000	0.42 ± 0.000	-0.15 ± 0.000	0.60 ± 0.000	0.26 ± 0.000	0.47 ± 0.000
	40	0.39 ± 0.000	0.43 ± 0.000	-0.14 ± 0.001	0.61 ± 0.000	0.26 ± 0.000	0.50 ± 0.001
	60	0.40 ± 0.001	0.43 ± 0.001	-0.12 ± 0.001	0.61 ± 0.000	0.27 ± 0.001	0.52 ± 0.001
	80	0.41 ± 0.001	0.45 ± 0.001	-0.10 ± 0.002	0.63 ± 0.001	0.28 ± 0.001	0.54 ± 0.002
	100	0.42 ± 0.002	0.45 ± 0.001	-0.09 ± 0.003	0.63 ± 0.001	0.28 ± 0.002	0.55 ± 0.001
	150	0.43 ± 0.003	0.46 ± 0.002	-0.08 ± 0.004	0.64 ± 0.001	0.29 ± 0.002	0.58 ± 0.003
	200	0.45 ± 0.004	0.49 ± 0.003	-0.03 ± 0.006	0.66 ± 0.002	0.32 ± 0.004	0.65 ± 0.004
	250	0.47 ± 0.006	0.49 ± 0.004	-0.01 ± 0.007	0.66 ± 0.002	0.32 ± 0.005	0.68 ± 0.003
	300	0.50 ± 0.004	0.51 ± 0.004	0.03 ± 0.008	0.67 ± 0.003	0.35 ± 0.006	0.72 ± 0.004
	400	0.53 ± 0.004	0.54 ± 0.003	0.09 ± 0.006	0.69 ± 0.002	0.40 ± 0.005	0.76 ± 0.007
	500	0.57 ± 0.013	0.55 ± 0.007	0.11 ± 0.015	0.69 ± 0.004	0.42 ± 0.012	0.79 ± 0.005

Table 3: Point estimates and standard errors of mean performance for models trained on the Hepatitis C Virus data. Row *Hybrid-B*: 0 Peptides corresponds to the "heterogeneous" case, with 0 organism- specific and 1,000 non-target pathogen peptides in the training set.

Training	Peptides	AUC	Bal. Accuracy	MCC	PPV	NPV	Sensitivity
OrgSpec	20	0.62 ± 0.001	0.57 ± 0.001	0.15 ± 0.001	0.43 ± 0.000	0.74 ± 0.001	0.78 ± 0.001
	40	0.65 ± 0.001	0.59 ± 0.001	0.18 ± 0.002	0.44 ± 0.001	0.75 ± 0.001	0.77 ± 0.003
	60	0.66 ± 0.002	0.61 ± 0.002	0.21 ± 0.003	0.46 ± 0.001	0.75 ± 0.002	0.74 ± 0.004
	80	0.68 ± 0.003	0.62 ± 0.002	0.24 ± 0.003	0.47 ± 0.001	0.77 ± 0.002	0.77 ± 0.004
	100	0.69 ± 0.002	0.63 ± 0.002	0.27 ± 0.003	0.48 ± 0.001	0.79 ± 0.002	0.78 ± 0.005
	150	0.70 ± 0.004	0.63 ± 0.004	0.27 ± 0.007	0.48 ± 0.003	0.78 ± 0.005	0.77 ± 0.008
	200	0.72 ± 0.007	0.65 ± 0.004	0.29 ± 0.008	0.50 ± 0.003	0.79 ± 0.006	0.74 ± 0.014
	250	0.72 ± 0.008	0.65 ± 0.005	0.30 ± 0.009	0.51 ± 0.006	0.79 ± 0.007	0.74 ± 0.014
	300	0.73 ± 0.006	0.66 ± 0.004	0.31 ± 0.008	0.51 ± 0.005	0.79 ± 0.006	0.75 ± 0.011
	400	0.75 ± 0.004	0.68 ± 0.008	0.34 ± 0.015	0.54 ± 0.009	0.80 ± 0.006	0.75 ± 0.007
	500	0.76 ± 0.003	0.69 ± 0.004	0.38 ± 0.009	0.56 ± 0.005	0.80 ± 0.005	0.73 ± 0.009
Hybrid-A	20	0.60 ± 0.001	0.56 ± 0.001	0.13 ± 0.001	0.44 ± 0.001	0.71 ± 0.001	0.68 ± 0.002
	40	0.63 ± 0.001	0.59 ± 0.001	0.17 ± 0.002	0.45 ± 0.001	0.72 ± 0.001	0.66 ± 0.004
	60	0.64 ± 0.002	0.60 ± 0.001	0.20 ± 0.003	0.47 ± 0.002	0.74 ± 0.002	0.69 ± 0.004
	80	0.66 ± 0.002	0.61 ± 0.001	0.22 ± 0.003	0.48 ± 0.001	0.75 ± 0.002	0.68 ± 0.006
	100	0.68 ± 0.003	0.62 ± 0.002	0.25 ± 0.003	0.49 ± 0.002	0.76 ± 0.003	0.71 ± 0.006
	150	0.68 ± 0.005	0.62 ± 0.004	0.24 ± 0.008	0.49 ± 0.005	0.75 ± 0.005	0.71 ± 0.008
	200	0.70 ± 0.007	0.63 ± 0.004	0.26 ± 0.008	0.50 ± 0.003	0.75 ± 0.005	0.67 ± 0.014
	250	0.70 ± 0.009	0.65 ± 0.006	0.29 ± 0.011	0.53 ± 0.008	0.76 ± 0.006	0.66 ± 0.013
	300	0.71 ± 0.004	0.66 ± 0.004	0.30 ± 0.009	0.53 ± 0.008	0.77 ± 0.003	0.70 ± 0.01
	400	0.73 ± 0.003	0.67 ± 0.006	0.33 ± 0.011	0.55 ± 0.004	0.77 ± 0.009	0.67 ± 0.023
	500	0.74 ± 0.004	0.68 ± 0.002	0.36 ± 0.007	0.58 ± 0.014	0.77 ± 0.006	0.65 ± 0.025
Hybrid-B	0	0.49 ± 0.001	0.48 ± 0.001	-0.04 ± 0.003	0.35 ± 0.002	0.61 ± 0.001	0.38 ± 0.004
	20	0.56 ± 0.001	0.55 ± 0.001	0.10 ± 0.001	0.42 ± 0.000	0.67 ± 0.001	0.55 ± 0.001
	40	0.58 ± 0.001	0.56 ± 0.001	0.11 ± 0.002	0.43 ± 0.001	0.68 ± 0.001	0.57 ± 0.003
	60	0.60 ± 0.002	0.57 ± 0.001	0.14 ± 0.003	0.44 ± 0.001	0.69 ± 0.002	0.58 ± 0.004
	80	0.63 ± 0.002	0.59 ± 0.002	0.18 ± 0.003	0.46 ± 0.002	0.71 ± 0.002	0.63 ± 0.005
	100	0.64 ± 0.003	0.60 ± 0.003	0.20 ± 0.005	0.47 ± 0.002	0.72 ± 0.003	0.63 ± 0.005
	150	0.66 ± 0.004	0.60 ± 0.003	0.19 ± 0.006	0.46 ± 0.002	0.73 ± 0.004	0.66 ± 0.009
	200	0.67 ± 0.007	0.63 ± 0.004	0.25 ± 0.007	0.51 ± 0.003	0.74 ± 0.005	0.63 ± 0.016
	250	0.69 ± 0.01	0.63 ± 0.005	0.26 ± 0.009	0.50 ± 0.002	0.75 ± 0.009	0.66 ± 0.022
	300	0.70 ± 0.008	0.65 ± 0.006	0.29 ± 0.012	0.53 ± 0.007	0.75 ± 0.006	0.66 ± 0.011
	400	0.73 ± 0.004	0.67 ± 0.004	0.33 ± 0.009	0.55 ± 0.009	0.78 ± 0.002	0.70 ± 0.009
	500	0.74 ± 0.004	0.68 ± 0.005	0.36 ± 0.011	0.58 ± 0.016	0.78 ± 0.004	0.67 ± 0.022