

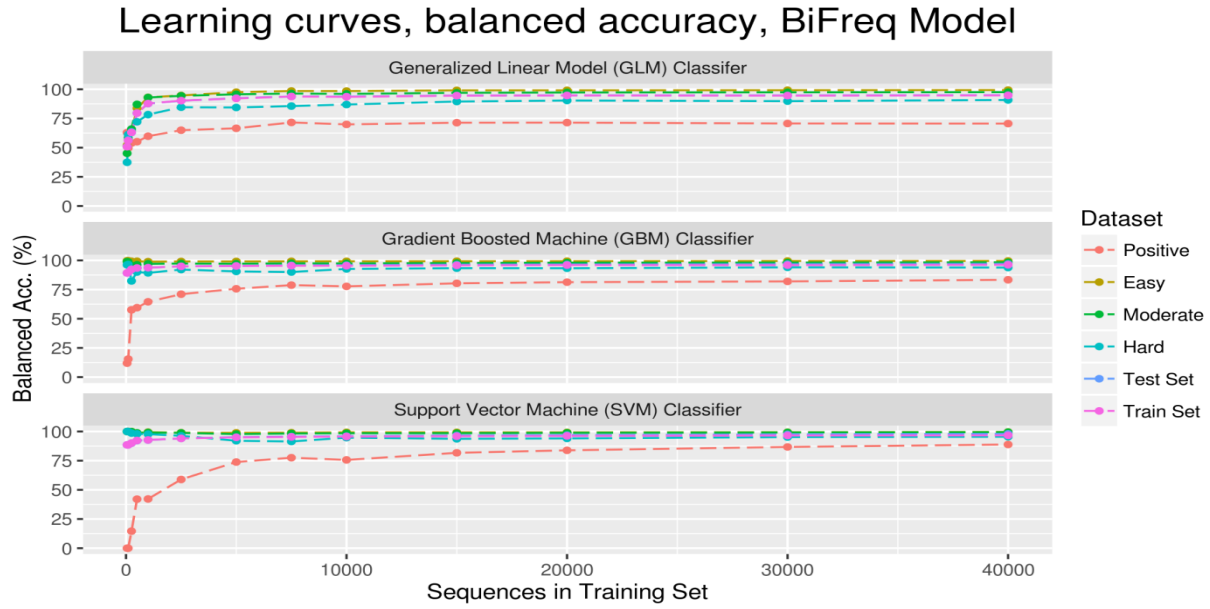
Supplementary Figure 1: Machine learning classifiers learning curves

Supplementary Figure 1 displays learning curves for the training of the learning classifiers using the different toxin prediction models. The models are described in Supplementary Note 1A. The learning curves compare the accuracy of prediction of the test set (y-axis) to the number of sequences in the training set (x-axis). The results are displayed for the Positive, Easy, Moderate and Hard datasets, as well as training and test datasets composed from the sum of the Positive, Easy, Moderate and Hard data.

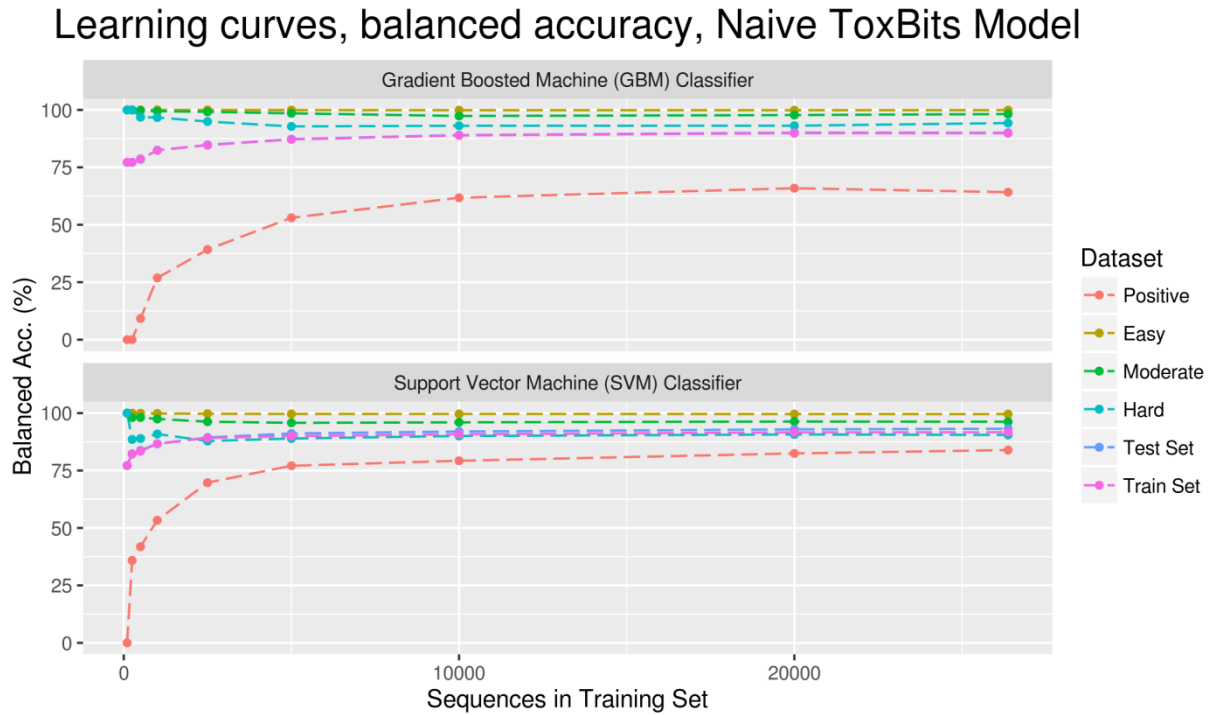
A: OneFreq (OF) model learning curves



B: BiFreq (BF) model learning curves

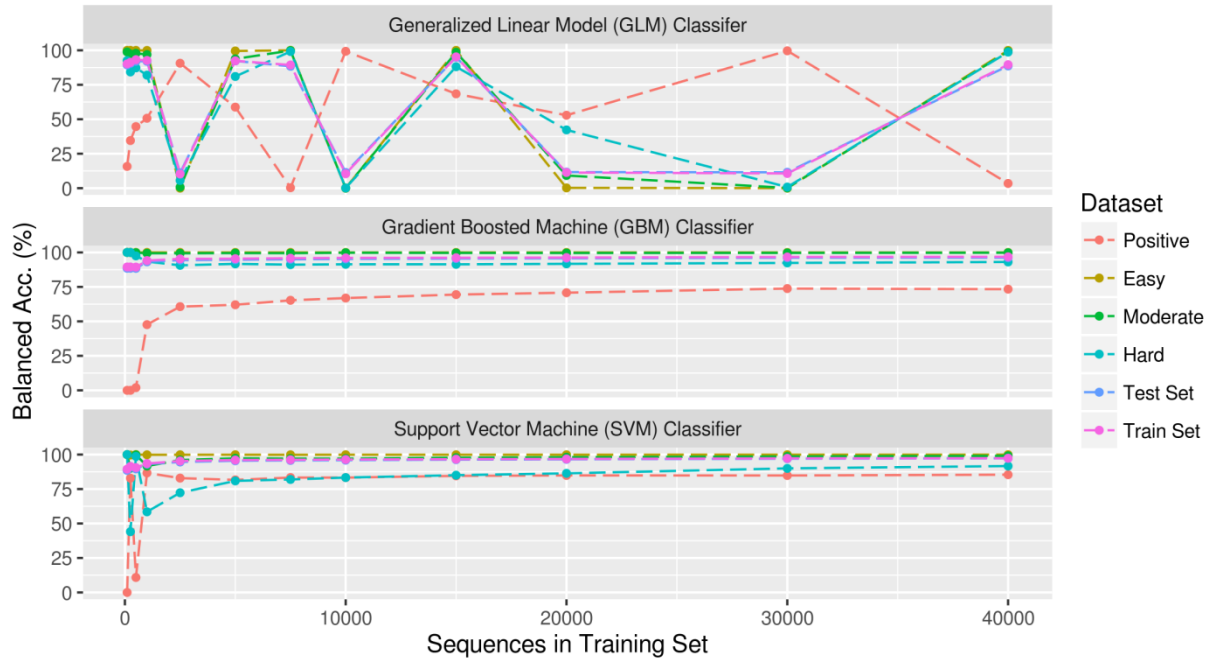


C: Naïve ToxBits (NTB) model learning curves



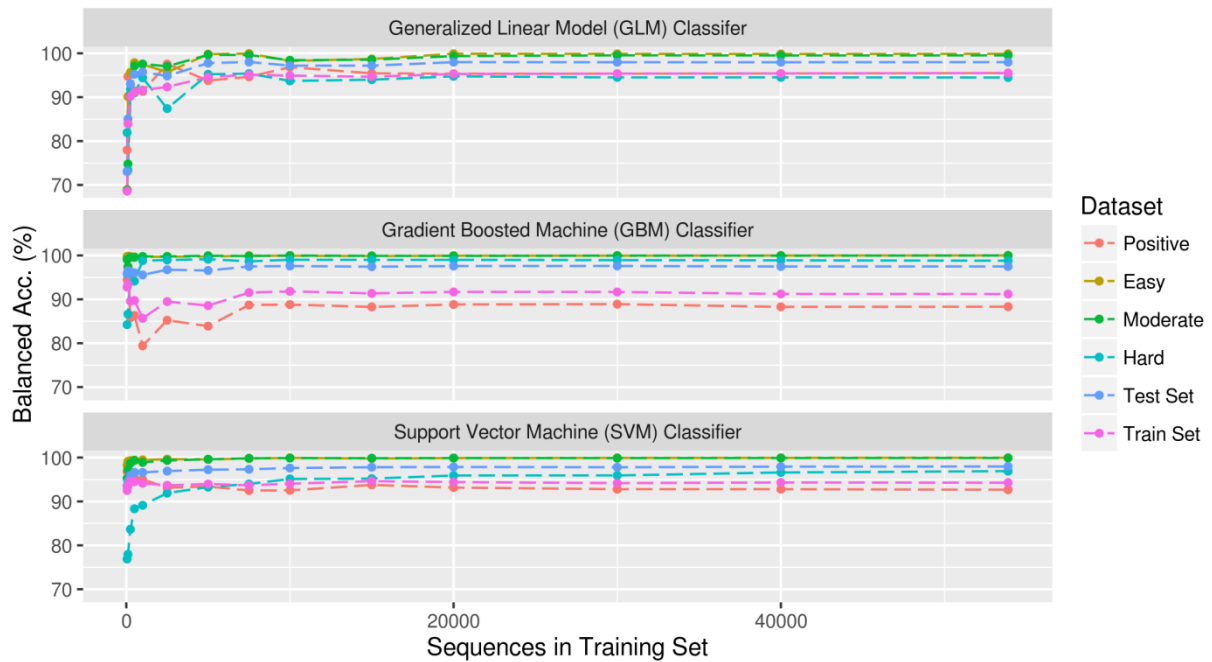
D: Scored ToxBits (STB) model learning curves

Learning curves, balanced accuracy, Scored ToxBits Model A



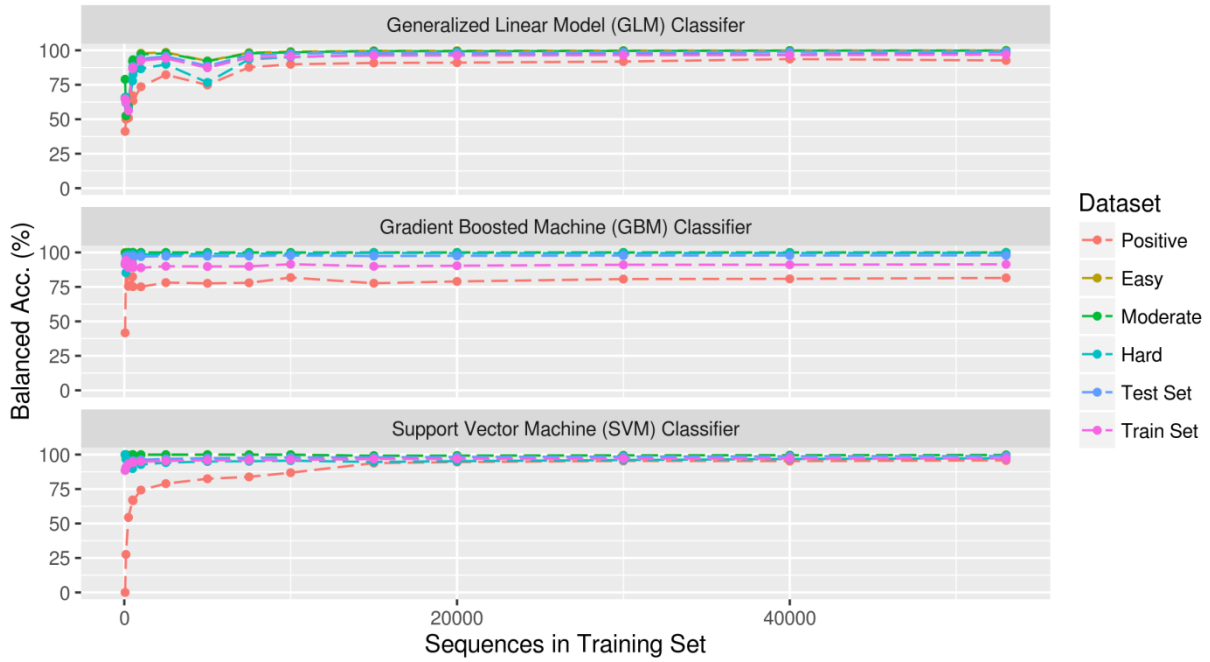
E: Tri BLAST 'simple' (TBS) model learning curves

Learning curves, balanced accuracy, Tri Blast Model



F: Tri BLAST 'enhanced' A (TBEa) model learning curves

Learning curves, balanced accuracy, Tri Blast Enhanced Model



G: Tri BLAST 'enhanced' B (TBEb) model learning curves

Learning curves, balanced accuracy, Tri Blast Enhanced Model B

