Some comments on the benchmark results that I uploaded:

* IEDB data comes from the select epitope sequences from which only the full sequence of the proteins were taken as data.
* The algpred2 data exists in 2 sets. For the ‘normal’ set “..algpred2..” the pre-splitted data was taken and used as given on the website of algpred2.0. The other set “..algpred2\_resplit..” was generated by concatenating the whole data set (their train and test datasets) and then splitted into 80/20 sets to train and test by myself. (I did this to test how much the test set and train set make a difference on the result.)
* FFNN is just feed forward neural network

If you have any questions feel free to ask me.

Rikard

(latest upload ProtBert embedding of algpred2 and ProtBert/ ESM2 embedding + DSSP features)