

Supplementary material 4

Definition of scoring indices used for evaluating the HMM. The closer to 1 the results, the better the method.

Accuracy

The accuracy (ACC) of the model that is a measure of the correctly identified classes over the total predictions:

$$ACC = \frac{(TP+TN)}{(TP+TN+FN+FP)}$$

True positive rate (TPR)

The sensitivity (or true positive rate(TPR)) of the model measures the proportion of positives that are correctly identified:

$$TPR = \frac{TP}{(TP+FN)}$$

True negative rate (TNR)

The specificity (or true negative rate(TNR)) of the model measures the proportion of negatives that are correctly identified:

$$TNR = \frac{TN}{(TN+FP)}$$

Matthew Correlation Coefficient (MCC)

The Matthew correlation coefficient (MCC) of the model is a particularly used for evaluating the binomial classifiers and it is a measure of the performance of the model that reduces the influence of the bias that could be present due to the different sizes of the classes analyzed:

$$MCC = \sqrt{\frac{(TP*TN)-(FP*FN)}{(TP+FP)*(TP+FN)*(TN+FP)*(TN+FN)}}$$

Root Mean Square Deviation (RMSD)

The Root Mean Square Deviation (RMSD) of atomic positions is the measure of the average distance between the atoms (usually the backbone atoms) of superimposed proteins:

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^n \delta_i^2}$$

where δ_i is the distance between atom i and either a reference structure or the mean position of the N equivalent atoms.