

UniProt, the largest public database of protein sequences

Phyre2 is the update of tools available on the web for the prediction and analysis of the structure, function and mutations of proteins

Also, an extended evaluation of methods for predicting protein function has been performed by conducting the second critical assessment of functional annotation (CAFA)

PROVEAN is a prediction tool for protein sequence from any organisms that has been presented as a web server that can predict the functional effect of single or multiple amino acid substitutions, insertions, and deletions

Oxygen binding proteins are ancient molecules that have probably evolved from enzymes that preserved the organism against the toxic oxygen

Oxypred is the first proposed tool for the classification and the prediction of oxygen binding proteins based on support vector machine

Recently, Random Forest machine learning tool has been proposed and achieved 89.22% of accuracy using amino acid composition [21].

Solving oxygen binding proteins prediction and classification problem aims at developing a model M that helps identifying the type of a protein and consequently its function.

RELU /one-hot encoding to perfect RELU /added a dropout regularization technique to prevent neural network from overfitting /droppout to prevent over-emulating/created 6 nodes in the output layer /Softmax function

UniProt database available in [21]. The datasets consist of protein sequences in the standard formats FASTA.

The numerical representation of the protein sequences was generated using Protr which is a comprehensive R package [32]. Then, the models have been trained and validated on Theano using Keras deep learning library in Python with Anaconda.

We used Keras which is a python library for deep learning that covers the efficient numerical computation libraries Theano, Tensorflow and CNTK to develop and validate our proposed models for classification of oxygen binding proteins.

We also used a logarithmic loss function which is called “categorical-crossentropy” for multiple classification and “binary-crossentropy” for binary classification in Keras.

we split our datasets in a way to get75% for training and 25% for testing purpose.

we plotted the area under the ROC (Receiver Operating Characteristic) curve, or AUC and the precision-recall curve for both cases.