Qeios

Peer Review

Review of: "Evaluation of Molecular Docking by Deep Learning and Random Forests: A Hybrid Approach Based on Pseudo-Convolutions"

A. V. Tuzikov[1]

1. United Institute of Informatics Problems, Minsk, Belarus

The article considers a problem of predicting the docking of two proteins based on their linear sequences. A proposed algorithm represents the sequences of two proteins as a numerical vector and uses a random forest classifier to predict possible interactions between the proteins.

The algorithm proposed in the article raises many questions. It is not very clear why a multi-step procedure of transforming the sequences of two proteins into a numerical vector is used. Why not construct a cooccurrence matrix directly based on the original sequences of two proteins with the same success?

How can the resulting numerical vector be related to the docking of two proteins? Actually, it does not carry any information about the structural complementarity of the two proteins under consideration and, accordingly, can hardly be a good characteristic for predicting protein docking.

The initial training set for constructing the classifier is too unbalanced in the number of positive and negative docking examples, and accordingly, it is advantageous for the classifier to predict the absence of docking. In this case, the probability of a prediction error is significantly reduced. From table 6, it is clear that the classifier does not correctly predict docking but is actually oriented towards predicting the absence of docking; see the values of the specificity measure and the proportion of FN cases.

I have serious doubts that the developed approach can have any competitive advantages in docking prediction compared to other approaches proposed in the literature.

I am sorry, but from my point of view, the paper should be rejected.

doi.org/10.32388/AX3O5K

Declarations

Potential competing interests: No potential competing interests to declare.