Qeios

Peer Review

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

Marcelo Fernandes[1]

1. Universidade Federal do Rio Grande do Norte, Brazil

The article proposes a hybrid architecture to predict molecular docking between protein pairs using pseudo-convolutions and Random Forests, eliminating the need for 3D structural modeling. The approach transforms protein sequences into co-occurrence matrices, which are converted into feature vectors and classified. The datasets used include Affinity Benchmark 3 (interacting proteins) and Negatome 2 (non-interacting proteins) for training and testing. Results achieved 94% accuracy, 93% AUC, and high sensitivity (94%), highlighting the method as an efficient and cost-effective alternative for molecular docking analysis and drug discovery.

Suggestions for Major Review

Test the model on external independent datasets to validate its generalization capabilities, ensuring the results do not balance current data.

Conduct a detailed analysis to confirm whether any proteins are shared between the Affinity Benchmark 3 and Negatome 2 datasets, as overlap could influence the results.

Provide a more explicit rationale for selecting Random Forest as the primary model compared to more advanced techniques such as Convolutional Neural Networks (CNNs) or Transformers.

The specificity of 78% suggests room for improvement. To better handle data imbalance, explore techniques such as class balancing (e.g., SMOTE) or hyperparameter tuning.

Include a direct comparison of the results with traditional 3D modeling methods and other machine-learning approaches for molecular docking, detailing relative advantages and limitations.

Declarations

Potential competing interests: No potential competing interests to declare.

doi.org/10.32388/B5MVZ1