**REINDEER: A Protein-Ligand Feature Generator Software for Machine Learning Algorithms**

Milad Rayka1\*

1. Unaffiliated

Corresponding author:

Milad Rayka, milad.rayka@yahoo.com

**Abstract**

**Keywords:** Cheminformatics, Protein-Ligand Complex, Software, Machine Learning, Feature Engineering

1- Introduction

In the past decade we witness a proliferation of data-driven approach for designing protein-ligand scoring function….

Classification to ML and DL-based…

Use new review paper

Indicate to number of examples

Emphasis of ML-based

Lack of package to feature engineering

Comparing to existence one, use table

Last paragraph of introduction

2- Theory

Occurrence of Interatomic Contact

Occurrence of interatomic contact (OIC) was introduced by Ballester et al. for developing the RF-Score scoring function [1]. In this technique, the authors represented a protein-ligand complex by counting the number of occurrences of a specific pair of protein and ligand atoms below a distance threshold. Nine elemental atom types (H, C, O, N, F, P, S, Cl, Br, and I) were allocated for protein and ligand. The following formula calculates the occurrence (1):

|  |  |
| --- | --- |
|  | (1) |

Where xi,j is the number of contacts between i and j atom types. k and l are protein and ligand atoms belonging to the i and j atom types. dkl is the Euclidean distance between k and l atoms, and Θ is the Heaviside step function that counts contacts below dcutoff=12 Å.

3- Method

4- Case Study

5- Conclusion

6- References

1. Ballester PJ, Mitchell JB. Bioinformatics 26, 9 (2010): 1169-1175.