**User Manual for nene**

Jurica Novak, Maria A. Grishina and Vladimir A. Potemkin

December 11, 2019

novaki@susu.ru, potemkinva@susu.ru

# Introduction

nene is computer program consisting of a series of machine learning algorithms designed to perform classification of enzymes according to their EC family and class based only on its radial distribution function (RDF). The code is written in Fortran 95, and compiled to be executed on Unix machines.

# Program description

The details of algorithm and its implementation can be found in

* Potemkin V, Grishina M. Grid-Based Technologies for In Silico Screening and Drug Design. Curr. Med. Chem. 2018; 25(29): 3526 - 3537.
* Jurica Novak, Maria A. Grishina and Vladimir A. Potemkin. The classification of enzymes using radial distribution function and artificial neural networks. Bioinfomatics 2020 (submitted)

**nene.exe** module builds a model for the classification. The usage is

nene.exe rdf.txt ec.txt

*rdf.txt* file is the file containing the radial distribution functions of the series of the enzymes. Each RDF is in one line, consisting of 1275 columns of numbers in f14.4 format.

*ec.txt* file is the file containing binary classification of the enzymes – 1. if it belongs to the selected class and 0. if it does not belong to the selected class. Point after 0 and 1 is mandatory. The ordering of the enzymes in rdf.txt and ec.txt files should be the same.

*rdf\_res.txt* file is an output file, with the correlation coefficient of the calculated and actual classification, standard deviation, cumulative classification error of the first and second kind and the quality of the classification in the first line. Starting from the second line are actual class, class predicted by the model probability to be assigned to class 1.

*rdf.rf* file is file with parameters of the neural network and discriminant analysis needed for predictions by *nene\_pred.exe* module.

**nene\_pred.exe** module is a prediction model. The usage is:

nene\_pred.exe rdf\_unknown.txt rdf.rf

*rdf\_unknown.txt* is file with the RDF of enzymes for which classification is going to be performed. The file format is analogous as *rdf.txt*.

test\_nn\_rdf\_pred.txt is the file with results. First number is classification result [enzyme is a member of the class (1.0) or it is not member of the class (0.0)], and the second is probability to be assigned to class 1.

# The used modules are for non-commercial use. We kindly ask to cite:

* Potemkin V, Grishina M. Grid-Based Technologies for In Silico Screening and Drug Design. Curr. Med. Chem. 2018; 25(29): 3526 - 3537.
* Jurica Novak, Maria A. Grishina and Vladimir A. Potemkin. The classification of enzymes using radial distribution function and artificial neural networks. Bioinfomatics 2020 (submitted)