Journal of Information and Computing Science
J. Info. Comput. Sci., Vol. 18, No. 1, pp. 27-50

DOI: 10.4208/JICS-2023-003
June 2023

Prediction of Properties of Anti-Breast Cancer Drugs Based on PSO-BP Neural Network and PSO-SVM

Meixian Xu[1], Yan Zheng[1,*], Yanju Li[1] and Weihao Wu[1]

[1] *College of Automobile and Traffic Engineering, Nanjing Forestry University, Nanjing 210044, China*

Abstract. The process of screening and developing new drugs through experiments is very slow and requires a lot of manpower and material resources, and the use of computer-aided prediction of the molecular properties of drugs can greatly save time and cost of drug development. Therefore, in order to enable anti-breast cancer candidate drugs to have good biological activity and ADMET properties for inhibiting ERα, the random forest classifier was first used for the collected 1 974 compounds to screen the top 20 molecular descriptors with the most significant effects on biological activity. Then a QSAR model was established using this and pIC_{50} value as characteristic data. The biological activity values of 50 new compounds were predicted via the PSO optimized BP neural network, with the model fit of 0.833 7 and the root mean square error of 0.731 5, which were more consistent with the actual values than the predicted results of the BP neural network. Subsequently, in order to improve the success rate of drug development, the ADMET classification prediction model was constructed using PSO to optimize the SVM based on the existing ADMET property data. The algorithm cross-validation CV accuracy rate reached 94.076 7%, and the prediction accuracy rates of the five index models were all above 79%. The results show that the proposed model has better prediction performance than the benchmark model, and the adopted prediction strategy is effective, which can provide reference for the discovery and development of anti-breast cancer drugs.

AMS subject classifications: 62M20, 92B20

Key words: Anti breast cancer drugs, Biological activity, ADMET properties, Particle Swarm Optimization (PSO), BP neural network, Support Vector Machines (SVM).

1 Introduction

Translated from *Journal of Nanjing University of Information Science & Technology*, 2023, 15(1): 51-65.
*Corresponding author. *Email addresses:* **xumeixian3210@163.com** (M. Xu), **ZhengYan3210@163.com** (Y. Zheng).

According to the 2018 cancer data report of the American Cancer Center, breast cancer is the most common malignant tumor in women in the world, and it seriously threatens the physical and mental health of women [1]. Breast cancer has become a worldwide health care problem, and treatment options need to be both selective and take into account the probability of effectiveness. To solve this problem, a large number of drug candidates have been studied and analyzed in the field of medicinal chemistry. The experimental results of estrogen receptor α subtype (ERα) gene deletion in mice show that ERα is considered to be an important target for the treatment of breast cancer, and compounds that can antagonize ERα activity may be candidates for the treatment of breast cancer.

Anti-breast cancer drug candidates need to have good biological activity from development to use, and their pharmacokinetic properties and safety should also meet the requirements of relevant policies and regulations. If only experimental methods are used to evaluate the biological activity, pharmacokinetic properties and safety of compounds, the time and cost will be immeasurable. Its pharmacokinetic properties and safety are collectively known as ADMET (Absorption, Distribution, Metabolism, Excretion, Toxicity) properties. Moreover, the data obtained from experimental animals does not fully coincide with clinical data, so it cannot meet the needs of modern drug research [2]. In order to save time and cost, research institutions usually choose to combine in vitro research techniques with computer computing models to establish compound activity prediction models and screen potential active compounds. That is, by collecting a series of compounds acting on ERα and their biological activity data, and selecting a series of molecular structure descriptors as independent variables, and the biological activity value of the compound as dependent variables, a quantitative structure-activity relationship (QSAR) model of the compound is constructed, and then the model is used to predict new compound molecules with better biological activity. Or to guide the structural optimization of existing active compounds. In addition to biological activity, pharmacokinetic properties and toxicity (ADMET) are also important factors in determining the success of drug development. No matter how good the activity of a compound is, if its ADMET properties are not good, such as difficult to be absorbed by the body, or the body metabolism is too fast, or has some toxicity, then it is still difficult to become a drug, so it also needs to optimize the ADMET properties.

In the case of the rapid increase in the number of drugs, the most economical and reasonable research method is to use computer-aided artificial intelligence algorithm to predict the biological activity of drugs and the properties of ADMET. Gu et al. [3] collected a large amount of drug ADMET data from multiple public databases and proposed to use graph neural network model for virtual screening of drug development after effective data cleaning. The research results showed that the model had good predictive performance and could be generalized. Considering the accuracy and fit of shallow and deep neural networks, Xie et al. [4] chose to