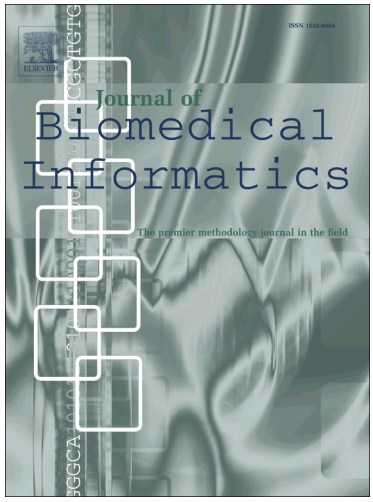

Classification of Toxicity Effects of Biotransformed Hepatic Drugs using Whale Optimized Support Vector Machines

Alaa Tharwat, Yasmine S. Moemen, Aboul Ella Hassanien

PII: S1532-0464(17)30051-5
DOI: http://dx.doi.org/10.1016/j.jbi.2017.03.002
Reference: YJBIN 2736

To appear in: *Journal of Biomedical Informatics*

Received Date: 7 September 2016
Accepted Date: 5 March 2017

Please cite this article as: Tharwat, A., Moemen, Y.S., Hassanien, A.E., Classification of Toxicity Effects of Biotransformed Hepatic Drugs using Whale Optimized Support Vector Machines, *Journal of Biomedical Informatics* (2017), doi: http://dx.doi.org/10.1016/j.jbi.2017.03.002

Classification of Toxicity Effects of Biotransformed Hepatic Drugs using Whale Optimized Support Vector Machines

Alaa Tharwat[a,b,1,*], Yasmine S. Moemen[b,c], Aboul Ella Hassanien[b,d]

[a] *Faculty of Engineering, Suez Canal University, Ismailia, Egypt*
[b] *Scientific Research Group in Egypt (SRGE), http://www.egyptscience.net*
[c] *Clinical Pathology Department, National Liver Institute, Menoufia University, Egypt*
[d] *Faculty of Computers and Information, Cairo University, Egypt*

Abstract

Measuring toxicity is an important step in drug development. Nevertheless, the current experimental methods used to estimate the drug toxicity are expensive and time-consuming, indicating that they are not suitable for large-scale evaluation of drug toxicity in the early stage of drug development. Hence, there is a high demand to develop computational models that can predict the drug toxicity risks. In this study, we used a dataset that consists of 553 drugs that biotransformed in liver. The toxic effects were calculated for the current data, namely, mutagenic, tumorigenic, irritant and reproductive effect. Each drug is represented by 31 chemical descriptors (features). The proposed model consists of three phases. In the first phase, the most discriminative subset of features is selected using rough set-based methods to reduce the classification time while improving the classification performance. In the second phase, different sampling methods such as Random Under-Sampling, Random Over-Sampling and Synthetic Minority Oversampling Technique (SMOTE), BorderLine SMOTE and Safe Level SMOTE are used to solve the problem of imbalanced dataset. In the third phase, the Support Vector Machines (SVM) classifier is used to classify

*Corresponding author
Email addresses: engalaatharwat@hotmail.com (Alaa Tharwat), yasmine_moemen@livereg.org (Yasmine S. Moemen), aboitcairo@gmail.com (Aboul Ella Hassanien)
[1] My present address is Faculty of Engineering, Suez Canal University, Ismailia, Egypt