

## **Pairwise Drug Interaction Prediction, Integrating, Clustering, and Classification**

P. N. Hameed\*

*Department of Computer Science, Faculty of Science, University of Ruhuna, Sri Lanka*

Investigating drug interactions is vital when administering multiple drugs for patients. However, experimentally-based drug interaction prediction consumes a large investment of money and time. Computational-based drug interaction prediction has shown significant benefits during the last two decades. Supervised and unsupervised machine learning approaches are frequently used to classify drug interactions based on drug characteristics. However, drug interactions cannot be classified only based on homogeneous properties as they have their limitations. Hence, investigating computational methods for heterogeneous data integration becomes necessary. Moreover, employing a representative training sample is crucial for obtaining a better classification of drug interactions. Though there are standard data on harmful drug interactions, there are no standard data for non-harmful drug interactions. Thus, investigating methods to find representative negatives is crucial. The proposed approach has two folds: (i) using an unsupervised two-tiered clustering approach for drug-pair clustering and (ii) using supervised classification for drug interaction classification. This study considered chemical, disease, protein, and side effects characteristics of drugs providing an opportunity to demonstrate drug characteristics from those four perspectives. The two-tiered clustering approach was used in the first fold that enables drug-pair clustering as well as heterogeneous data integration. The clustered result can be used to infer plausible negatives for drug interaction classification. In the second phase, binary classifiers such as Support Vector Machine, Logistic Regression, and Random Forest can be used. Applying an ensemble learning model integrating with the results of multiple classifiers could further improve the clinical significance of the predicted drug interactions.

*Keywords:* Drug interactions, Heterogeneous data integration