Development of a cardiotoxicity prediction scheme using Sub-Graphed Transformer Neural Network

Name: WANG HUIJIA

Laboratory's name: Computational Systems Biology

Supervisor's name: Prof. Shigehiko Kanaya, Prof. Keiichi Yasumoto, Assoc Prof. MD.Altaf-Ul-Amin, Assoc Prof. Naoaki Ono, Visiting Assoc Prof. Alex Ming Huang

Abstract:

This study introduces a Sub-graphs-based Transformer Neural Network (GSTN) model, leveraging heterogeneous graphs to predict cardiotoxicity in pharmaceutical compounds. By defining distinct sub-graphs and employing meta-paths, the model effectively aggregates and propagates intricate molecular structural information, crucial for identifying cardiotoxicity. We integrated the domain knowledge of cardiotoxicity, and addressed the challenges of non-uniform labeling standards by evaluating the impact of various thresholds on model performance, ultimately selecting a dual threshold strategy. These methodologies employed in GSTN markedly outperform traditional GCN models, achieving an impressive Accuracy of 90.5%, a Precision of 90.4%, a Recall of 90.4%, an AUC of 90.4%, and an F1-score of 90.0%. Additionally, we validated 56 FDA-approved drugs, assessing their potential cardiotoxicity to ensure safety in medical applications. Furthermore, to counter the "black-box" nature of deep learning models, we incorporated a visualization technique that assigns different weights to each atom, thus illuminating molecular regions contributing to toxicity. This advancement not only improves interpretability but also guides researchers in avoiding toxic molecular structures. In summary, this paper contributes to the field of drug safety analysis by developing a robust, interpretable, and efficient GSTN model.