INFORMATION OF THE DOCTORAL THESIS

Thesis title:

RESEARCH AND PROPOSE COMPUTATIONAL METHODS TO PREDICT DRUG RESPONSE IN THE TREATMENT

Speciality: **Information Systems**
Code: **9.48.01.04**
PhD. Candidate: **Nguyen Thi Thu Giang**
Scientific Supervisors:

1. **Assoc. Prof. Le Duc Hau, PhD**
2. **Assoc. Prof. Nguyen Trong Khanh, PhD**

Training institution: Posts and Telecommunications Institute of Technology

NEW FINDINGS OF THE THESIS:

Proposes for drug response prediction

(1) Propose a novel method for representing of molecular graph data of drug – GraphDRP. In this method, drugs are represented in molecular graphs as natural forms directly capturing the features of nodes as atoms and the bonds among them; meanwhile, cell lines were depicted as binary vectors of genomic aberrations. Representative features of drugs and cell lines were learned by convolution layers and then combined to represent drug-cell line pairs. Four variants of graph neural networks (GCN, GAT, GIN, GCN-GAT) were used for learning the features of drugs. GraphDRP has demonstrated the effectiveness of the proposed model in predicting drug response.

(2) *Propose a novel method for integrating molecular graph data of drugs and multiple-omics data of cell lines - GraOmicDRP*: This approach aims to integrate the representation of multiple-omics data of cell line and graph representation of drug to predict drug response. In this approach, drugs are represented as graphs of bonds between atoms, while cell lines are depicted using not only genomic data but also transcriptomic and epigenomic data. Graph Isomorphism Network (GIN) block and one-dimensional convolutional neural network (CNN1D) ones are used to learn the representations of drugs and cell lines, respectively. The results show that the GraOmicDRP method not only demonstrates superior capability in integrating multi-omics data but also helps identify meaningful features that significantly contribute to the prediction process.

Proposes for drug synergy prediction

(3) Propose a method for integrating multiple-omics data of cell lines and molecular graph data of drugs for drug synergy prediction – GraOmicSynergy: This approach integrates multiple-omics data representations of cell lines with graph representations of drug pairs to predict drug synergy. In this method, representations of drug pairs are learned using Graph Isomorphism Networks (GIN) and synthesized through an attention mechanism that simulates drug testing and drug synergy measuring more naturally; meanwhile, cell lines are represented by different -omics data learned through CNN1D networks. GraOmicSynergy effectively indicates that the integration of multi-omics data and drug pair representation using attention mechanisms for predicting drug synergy.

(4) Propose a method for integrating multiple-omics and topological information of PPI network for drug synergy prediction - AE-XGBSynergy: This approach integrates multi-omics cell line data and drug-cell line features derived from the topological characteristics of the protein-protein interaction (PPI) network to predict drug synergy. AE-XGBSynergy leverages drug-protein and cell line-protein interactions to construct a multi-layer graph network based on structural similarities and contextual relationships between nodes, then extracts features from both the cell line and drug pair. These features are concatenated with high-dimensional omics data from the cell line (genomic and epigenomic), which are extracted using a pre-trained

encoder. The combined features are then used as input for an extreme gradient-boosting algorithm to predict drug synergies. Experimental results demonstrate that the proposed model excels at integrating multi-omics data for drug synergy prediction.

APPLICATIONS, PRACTICAL APPLICABILITY AND FURTHER RESEARCH DIRECTIONS:

The results derived from the thesis can be used in analyzing, evaluating, and improving the performance of drug response prediction models in disease treatment, thereby contributing to the effectiveness of preclinical and clinical research in precision medicine. These findings can also be utilized in teaching and research at universities. The research directions can be developed from this thesis including modeling drug representation data by integrating physicochemical properties and 3D structures; incorporating additional multi-omics data such as proteomics and metabolomics; and enhancing computational methods to improve the learning of drug and cell line representations such as molecular pretraining models, transformers, and graphformer … for drug response prediction.

Confirmation of representative supervisors		**PhD candidate**
Assoc.Prof. Lê Đức Hậu	**Assoc.Prof. Nguyễn Trọng Khánh**	**Nguyễn Thị Thu Giang**