

TRANG THÔNG TIN LUẬN ÁN TIẾN SĨ

Tên đề tài luận án tiến sĩ:

NGHIÊN CỨU ỨNG DỤNG VÀ ĐỀ XUẤT CÁC PHƯƠNG PHÁP TÍNH TOÁN ĐỂ DỰ ĐOÁN ĐÁP ỨNG THUỐC TRONG ĐIỀU TRỊ BỆNH

Chuyên ngành: **Hệ thống thông tin**

Mã số chuyên ngành: **9.48.01.04**

Nghiên cứu sinh: **Nguyễn Thị Thu Giang**

Hướng dẫn khoa học:

1. PGS.TS. Lê Đức Hậu.

2. PGS. TS. Nguyễn Trọng Khánh

Cơ sở đào tạo: Học viện Công nghệ Bưu chính Viễn thông

NHỮNG KẾT QUẢ MỚI CỦA LUẬN ÁN:

Đề xuất các giải pháp để dự đoán đáp ứng đơn thuốc

(1) *Đề xuất giải pháp học dữ liệu biểu diễn đồ thị của phân tử thuốc – GraphDRP.* Giải pháp áp dụng cách biểu diễn dữ liệu thuốc một cách tự nhiên dưới dạng đồ thị phân tử với các đỉnh là các nguyên tố hóa học, cạnh là các liên kết giữa chúng. Các đặc trưng ẩn của phân tử thuốc được học thông qua một số biến thể của mạng nơ-ron đồ thị (GCN, GAT, GIN, GCN-GAT), trong khi các dòng tế bào được biểu diễn dưới dạng các vectơ nhị phân mô tả thông tin đột biến gen di truyền (genomics), các đặc trưng biểu diễn được học thông qua các lớp tích chập một chiều (CNN1D). Các đặc trưng biểu diễn cho thuốc và dòng tế bào sau đó được kết hợp thành các biểu diễn đặc trưng cho từng cặp tương tác thuốc - dòng tế bào để dự đoán đáp ứng thuốc. Giải pháp này cho thấy hiệu quả của việc cải tiến cách học biểu diễn dữ liệu đồ thị phân tử thuốc trong dự đoán đáp ứng thuốc năng dự đoán tốt hơn so với phương pháp không áp dụng dạng biểu diễn đồ thị trên các kịch bản thử nghiệm.

(2) *Đề xuất giải pháp tích hợp đa dữ liệu -omics và dữ liệu đồ thị phân tử thuốc - GraOmicDRP:* Giải pháp thực hiện tích hợp đa dạng các dữ liệu biểu diễn các dòng tế bào không chỉ là dữ liệu gen di truyền (genomics) mà còn là dữ liệu methyl hóa (epigenomics), dữ liệu biểu hiện gen (transcriptomics) kết hợp với dữ liệu biểu diễn đồ thị phân tử thuốc để dự đoán đáp ứng thuốc. Trong đó, các biểu diễn đặc trưng phân tử thuốc được học dựa trên mạng nơ-ron đồ thị đẳng cấu (GIN) và dữ liệu biểu diễn dòng tế bào được tổng hợp từ các dữ liệu -omics khác nhau được trích xuất thông qua mạng CNN1D. GraOmicDRP đã cho thấy khả năng hiệu quả của việc tích hợp đa dữ liệu -omics trên các kịch bản thử nghiệm. Giải pháp tích hợp đa dữ liệu -omics này không chỉ cho thấy hiệu quả của việc tích hợp đa dữ liệu -omics trong dự đoán mà còn giúp xác định được dữ liệu có ý nghĩa trong mô hình dự đoán đáp ứng thuốc.

Đề xuất các giải pháp tích hợp dữ liệu để dự đoán đáp ứng đa thuốc

(3) *Đề xuất giải pháp học các biểu diễn đồ thị của đa phân tử thuốc và tích hợp đa dữ liệu -omics – GraOmicSynergy:* Giải pháp thực hiện tích hợp đa dữ liệu -omics như dữ liệu gen di truyền (genomics) methyl hóa (epigenomics), biểu hiện gen (transcriptomics) kết hợp với dữ liệu biểu diễn thuốc dưới dạng đồ thị phân tử để dự đoán đáp ứng đa thuốc. Trong đó, cặp thuốc được học thông qua mạng nơ-ron đồ thị đẳng cấu (GIN), và được tổng hợp thông qua một mô-đun cơ chế chú ý mô phỏng quá trình thử thuốc và đo độ kết hợp thuốc một cách tự nhiên hơn, trong khi đó dòng tế bào được tổng hợp dựa trên các dữ liệu -omics khác nhau được học thông qua các mạng CNN1D. GraOmicSynergy cho thấy hiệu quả của việc tích hợp đa dữ liệu -omics và dữ liệu biểu diễn tổng hợp thuốc dưới dạng đồ thị cơ chế chú ý trong việc dự đoán đáp ứng đa thuốc.

(4) Đề xuất giải pháp tích hợp đa dữ liệu -omics và thông tin mạng sinh học - AE-XGBSynergy: Giải pháp tích hợp đa dữ liệu -omics của dòng tế bào được trích xuất thông qua bộ mã hóa pre-trained encoder kết hợp với dữ liệu biểu diễn thuốc và dòng tế bào được trích xuất thông qua thông tin cấu trúc mạng tương tác protein (PPI) để dự đoán phân loại đáp ứng đa thuốc. AE-XGBSynergy khai thác các mối quan hệ thuốc – đích (drug-protein), dòng tế bào – protein, xây dựng mạng đồ thị đa tầng để trích xuất đặc điểm tương đồng về cấu trúc và tạo ra bối cảnh cấu trúc cho các nút và trích xuất đặc trưng của dòng tế bào và thuốc để tích hợp với các biểu diễn dòng tế bào như dữ liệu methyl hóa, dữ liệu hệ gen di truyền để dự đoán đáp ứng đa thuốc. AE-XGBSynergy cho thấy hướng tiềm năng trong việc tích hợp đa dữ liệu -omics với thông tin mạng tương tác protein để dự đoán sự kết hợp thuốc.

CÁC ỨNG DỤNG, KHẢ NĂNG ỨNG DỤNG TRONG THỰC TIỄN HOẶC NHỮNG VẤN ĐỀ CÒN BỎ NGỎ CẦN TIẾP TỤC NGHIÊN CỨU:

Các kết quả nghiên cứu của luận án có thể sử dụng trong việc phân tích, đánh giá và cải thiện hiệu năng của các mô hình dự đoán đáp ứng thuốc trong điều trị bệnh, góp phần nâng cao hiệu quả trong nghiên cứu điều trị tiền lâm sàng và lâm sàng trong y học chính xác. Các phân tích và kết quả này cũng có thể được sử dụng trong giảng dạy và nghiên cứu tại các trường đại học. Các hướng phát triển tiếp theo của luận án bao gồm: mô hình hóa dữ liệu biểu diễn thuốc tích hợp đặc trưng lý hóa, cấu trúc 3D; mô hình hóa dữ liệu tích hợp đa dữ liệu -omics khác như proteomics, metabomics; áp dụng các mô hình tính toán tiên tiến như: molecular-pretrain model, transformer, graphformer cho bài toán dự đoán đáp ứng thuốc.

**Xác nhận của đại diện tập thể
Người hướng dẫn khoa học**

Nghiên cứu sinh

PGS.TS. Lê Đức Hậu

PGS.TS. Nguyễn Trọng Khánh Nguyễn Thị Thu Giang