

## APPLICATION OF MACHINE LEARNING IN SCREENING FOR TYPE 2 ANTI-DIABETES NATURAL COMPOUNDS

Pham Vu Hoang Phuong<sup>1</sup>, Tran Huu Thang<sup>1</sup>, Tran Thi Khanh Linh<sup>1</sup>,  
Vu Hoang Quynh Anh<sup>1</sup>, Nguyen Thi Kim Nhung<sup>1,2</sup>, Nguyen Minh Nam<sup>1,2\*</sup>

<sup>1</sup>University of Health Sciences, Vietnam National University at Ho Chi Minh City - Hai Thuong Lan Ong Street, Ho Chi Minh City National University Urban Area, Dong Hoa Ward, Di An City, Binh Duong Province, Vietnam

<sup>2</sup>Center for Genetics and Reproductive Health (CGRH), University of Health Sciences, Vietnam National University at Ho Chi Minh City - Hai Thuong Lan Ong Street, Ho Chi Minh City National University Urban Area, Dong Hoa Ward, Di An City, Binh Duong Province, Vietnam

Received: 30/10/2024

Revised: 16/11/2024; Accepted: 26/11/2024

### ABSTRACT

**Objective:** Building and developing a machine learning model to predict natural compounds in the treatment of type 2 diabetes mellitus (T2DM), evaluating the multi-target impact ability of screened compounds.

**Materials and Methods:** The study uses ImRMR and XGBoost algorithms to build a machine learning model that predicts natural compounds with anti T2DM activity. Combined with Molecular Docking on AutoDock Vina, the study evaluates the multi-target binding between screened compounds and potential impact targets of T2DM.

**Results:** The DiMeNP model with an accuracy of 0.917 and an AUC of 0.962 was applied to screen for more than 13,000 natural compounds and predicted 11 flavonoid compounds and 3 lignan compounds. The compounds met the criteria of bioavailability and drug similarity, with a prediction probability of  $> 0.9$  and all of them were well aligned on 5 potential targets in the treatment of diabetes including GLP1, DPP4, PPAR- $\gamma$ ,  $\alpha$ -glucosidase, and  $\alpha$ -amylase.

**Conclusion:** Applying machine learning methods combined with molecular binding, the research and construction of the DiMeNP model is capable of screening compounds with multi-target impact activity on T2DM.

**Keywords:** Type 2 Diabetes, Natural Compounds, Machine Learning, Molecular Binding.

---

\*Corresponding author

**Email:** nmnam@uhsvnu.edu.vn **Phone:** (+84) 904972804 **Https://doi.org/10.52163/yhc.v65iCD12.1846**



# ỨNG DỤNG HỌC MÁY TRONG SÀNG LỌC HỢP CHẤT TỰ NHIÊN KHÁNG ĐÁI THÁO ĐƯỜNG TYP 2

Phạm Vũ Hoàng Phương<sup>1</sup>, Trần Hữu Thắng<sup>1</sup>, Trần Thị Khánh Linh<sup>1</sup>,  
Vũ Hoàng Quỳnh Anh<sup>1</sup>, Nguyễn Thị Kim Nhung<sup>1,2</sup>, Nguyễn Minh Nam<sup>1,2\*</sup>

<sup>1</sup>Trường Đại học Khoa học Sức khỏe, Đại học Quốc gia Thành phố Hồ Chí Minh - Đường Hải Thượng Lãn Ông,  
Khu đô thị Đại học Quốc gia Thành phố Hồ Chí Minh, P. Đông Hòa, Tp. Dĩ An, Tỉnh Bình Dương, Việt Nam  
<sup>2</sup>Trung tâm nghiên cứu di truyền và sức khỏe sinh sản, Trường Đại học Khoa học Sức khỏe, Đại học Quốc gia Thành phố  
Hồ Chí Minh - Đường Hải Thượng Lãn Ông, Khu đô thị Đại học Quốc gia Thành phố Hồ Chí Minh,  
P. Đông Hòa, Tp. Dĩ An, Tỉnh Bình Dương, Việt Nam

Ngày nhận bài: 30/10/2024

Chỉnh sửa ngày: 16/11/2024; Ngày duyệt đăng: 26/11/2024

## TÓM TẮT

**Mục tiêu:** Xây dựng và phát triển mô hình Học máy dự đoán các hợp chất tự nhiên trong điều trị Đái tháo đường typ 2 (ĐTĐT2), Đánh giá khả năng tác động đa mục tiêu của các hợp chất sàng lọc được.

**Phương pháp nghiên cứu:** Nghiên cứu sử dụng thuật toán ImRMR và XGBoost để xây dựng mô hình Học máy dự đoán các hợp chất tự nhiên có hoạt tính kháng ĐTĐT2. Kết hợp Gắn kết phân tử trên AutoDock Vina, nghiên cứu đánh giá khả năng gắn kết trên đa mục tiêu giữa các hợp chất sàng lọc được với các đích tác động tiềm năng của ĐTĐT2.

**Kết quả:** Mô hình DiMeNP với độ chính xác là 0,917 và AUC là 0,962 đã được áp dụng để sàng lọc hơn 13,000 hợp chất tự nhiên và dự đoán được 11 hợp chất flavonoid và 3 hợp chất lignan. Các hợp chất đáp ứng các tiêu chí về sinh khả dụng và tính giống thuốc, với xác suất dự đoán > 0,9 và đều gắn kết tốt trên 5 mục tiêu tiềm năng trong điều trị đái tháo đường bao gồm GLP1, DPP4, PPAR- $\gamma$ ,  $\alpha$ -glucosidase và  $\alpha$ -amylase.

**Kết luận:** Ứng dụng phương pháp học máy kết hợp gắn kết phân tử, nghiên cứu xây dựng mô hình DiMeNP có khả năng sàng lọc được các hợp chất có hoạt tính tác động đa mục tiêu trên ĐTĐT2.

**Từ khóa:** Đái tháo đường typ 2, Hợp chất tự nhiên, Học máy, Gắn kết phân tử.

## 1. ĐẶT VẤN ĐỀ

Đái tháo đường typ 2 (ĐTĐT2) là một bệnh rối loạn chuyển hóa mạn tính toàn cầu với nhiều biến chứng nguy hiểm [1]. Hiện nay, bệnh ĐTĐT2 vẫn chưa có phương pháp điều trị triệt để và để kiểm soát đường huyết cũng như biến chứng, bệnh nhân phải dùng thuốc cả đời [2]. Về lâu dài, sự phối hợp thuốc là cần thiết khi các thuốc hiện tại chủ yếu tập trung trên những mục tiêu đơn lẻ trong con đường bệnh sinh phức tạp của đái tháo đường (ĐTĐ) [3]. Do đó, bệnh nhân phải đối diện với các nguy cơ về tác dụng phụ, tương tác thuốc cũng như gánh nặng kinh tế kéo dài [4]. Gần đây, việc nghiên cứu các phương pháp điều trị mới có khả năng tác động đa mục tiêu đang trở thành hướng phát triển tiềm năng cho các bệnh lý chuyển hóa phức tạp như ĐTĐ [5]. Tận dụng các cơ sở dữ liệu mở lớn, nghiên cứu xây dựng mô hình Học máy để sàng lọc các hợp chất tự nhiên có tác

động đa mục tiêu trên ĐTĐT2. Bên cạnh đó, nghiên cứu cũng sử dụng phương pháp Gắn kết Phân tử để làm rõ cơ chế hoạt động cũng như đánh giá khả năng tác động đa mục tiêu của các hợp chất đã được sàng lọc thông qua mô hình học máy.

## 2. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

### 2.1. Thiết kế nghiên cứu

Nghiên cứu áp dụng phương pháp Học máy và Gắn kết phân tử.

### 2.2. Địa điểm và thời gian nghiên cứu

Nghiên cứu được thực hiện trên máy vi tính HP Pavilion

\*Tác giả liên hệ

Email: nmnam@uhsvnu.edu.vn Điện thoại: (+84) 904972804 <https://doi.org/10.52163/yhc.v65iCD12.1846>

15-cs3012TU i5 (8QP30PA) từ ngày 15/12/2023 đến ngày 21/09/2024.

### 2.3. Đối tượng nghiên cứu

Đối tượng nghiên cứu gồm: các thuốc điều trị ĐTDĐ đang được FDA chấp thuận hoặc đang được nghiên cứu trong các giai đoạn lâm sàng, thuốc không có khả năng điều trị ĐTDĐ (thuốc có các chỉ định khác), dược liệu và các hợp chất tự nhiên trong dược liệu, đích tác động tiềm năng trên ĐTDĐ2. Tiêu chí lựa chọn thuốc có hoạt tính (TD) gồm các thuốc điều trị ĐTDĐ đã được FDA phê duyệt hoặc đang trong giai đoạn nghiên cứu lâm sàng (các pha đầu 1, pha 1, pha 2, pha 3 và pha 4). Tiêu chí loại trừ thuốc không có hoạt tính (TÁ) gồm: thuốc điều trị ĐTDĐ, thuốc điều trị béo phì, thuốc cản quang và thuốc dùng ngoài. Tiêu chí loại trừ chung cho tất cả các thuốc gồm thuốc đã ngưng sử dụng, thuốc trùng thành phần hoạt chất và thuốc trùng PubChem ID.

### 2.4. Cơ mẫu, chọn mẫu

Nghiên cứu thu thập dữ liệu về thuốc, dược liệu, hợp chất tự nhiên từ dược liệu và các đích tác động của ĐTDĐ2 từ tháng 3/2024 - 9/2024.

### 2.5. Nội dung nghiên cứu

- Xây dựng mô hình học máy dự đoán khả năng kháng ĐTDĐ2 của các hợp chất tự nhiên.

- Sàng lọc hợp chất tự nhiên có khả năng kháng ĐTDĐ2.

- Gắn kết phân tử giữa các hợp chất được dự đoán kháng ĐTDĐ2 thỏa các tiêu chí về sinh khả dụng và tính giống thuốc với các đích tác động tiềm năng trên ĐTDĐ2.

### 2.6. Kỹ thuật, công cụ và quy trình thu thập số liệu

- Thu thập và xử lý dữ liệu: Dữ liệu về thuốc được thu thập từ tổng quan tài liệu và cơ sở dữ liệu ChEMBL. Các bộ lọc tìm kiếm trên ChEMBL bao gồm: nhóm phân tử nhỏ, không vi phạm định luật 5 Lipinski, trọng lượng phân tử trong khoảng 100 - 500, và AlogP trong khoảng 0 - 5. Dữ liệu sau đó được cân bằng hóa bằng cách chia nhỏ ngẫu nhiên thành các tập dữ liệu con dựa trên số lượng thuốc trong từng nhóm và giai đoạn lâm sàng. Các tập dữ liệu con được phân chia thành tỷ lệ 80:20, tương ứng với dữ liệu huấn luyện và kiểm tra. Dữ liệu về hợp chất tự nhiên được tải về dưới định dạng tệp ".sdf" từ cơ sở dữ liệu CMAUP (cập nhật lần cuối vào tháng 3/2024), cơ sở dữ liệu DiaNatDB (cập nhật lần cuối vào 2021) và cơ sở dữ liệu Vietherb. Các hợp chất từ Vietherb được sàng lọc tiêu chí về tính giống thuốc (Drug-likeness) và sinh khả dụng (Bioavailability) bằng công cụ tìm kiếm nâng cao trong cơ sở dữ liệu Coconut.

- Chuyển đổi cấu trúc SMILES thành mã nhị phân: Để chuyển đổi cấu trúc thuốc và hợp chất chưa có dữ liệu, Python được sử dụng từ nền tảng Kaggle để truy xuất PubChemID và Canonical SMILES. Các hợp chất này được chuẩn hoá lại bằng công cụ RDKit, sau đó mã hóa thành dạng biến nhị phân theo 4 loại dấu vân tay phân tử: Molecular Access System (MACCS), RDK,

Extended Connectivity Fingerprint 4 (ECFP4) và Functional Class Fingerprint 4 (FCFP4). Các dấu vân tay phân tử này được tổ hợp với nhau để tạo ra 15 tổ hợp dấu vân tay phân tử. Mỗi tổ hợp sau đó được phân tích sử dụng thuật toán ImRMR để sàng lọc và trích chọn ra các đặc trưng quan trọng nhất trong từng tổ hợp.

$$C_4^1 + C_4^2 + C_4^3 + C_4^4 = 15$$

- Xây dựng mô hình dự đoán các hợp chất tự nhiên có hoạt tính kháng đái tháo đường - Mô hình DiMeNP: Các tổ hợp dấu vân tay phân tử, sau khi được trích chọn đặc trưng quan trọng, được sử dụng để xây dựng mô hình dự đoán bằng thuật toán XGBoost. Mô hình có hiệu quả dự đoán tốt nhất được lựa chọn làm mô hình cuối cùng và được gọi là DiMeNP. Hiệu suất của mô hình được đánh giá dựa trên các chỉ số sau: độ nhạy (sensitivity), độ đặc hiệu (specificity), độ chính xác (accuracy), độ tinh xác (precision), điểm F1 (F1 score) và hệ số tương quan Matthews (MCC). Ngoài ra, mô hình cũng được đánh giá bằng cách tính diện tích dưới đường cong ROC (AUC) và phân tích ma trận nhầm lẫn (confusion matrix) để đánh giá toàn diện hiệu suất dự đoán của mô hình

Sau khi chọn được mô hình có hiệu quả dự đoán tốt nhất, mô hình này được đánh giá trên tập dữ liệu kiểm tra độc lập từ cơ sở dữ liệu DiaNatDB, bao gồm 336 hợp chất tự nhiên đã được nghiên cứu *in vitro* hoặc *in vivo* về hoạt tính điều trị ĐTDĐ. Mục tiêu của việc đánh giá này là xác nhận xem các hợp chất mà mô hình dự đoán có hoạt tính có tương thích với kết quả của các nghiên cứu đã công bố trước đó hay không. Việc kiểm tra này giúp đảm bảo rằng mô hình DiMeNP không chỉ có hiệu suất tốt trên tập dữ liệu huấn luyện mà còn có khả năng áp dụng rộng rãi và đáng tin cậy trên các hợp chất tự nhiên đã được xác thực trong thực nghiệm.

- Sàng lọc hợp chất từ cơ sở dữ liệu CMAUP và Vietherb: Các hợp chất từ cơ sở dữ liệu CMAUP và Vietherb được dự đoán hoạt tính kháng ĐTDĐ bằng mô hình DiMeNP với giá trị xác suất > 0,9 được lựa chọn và sau đó sau đó được sàng lọc hai tiêu chí Sinh khả dụng > 0,3 và Tính giống thuốc > 0,18 bằng các công cụ tương ứng lần lượt là SwissADME và Molsoft.

- Thu thập đích tác động và gắn kết phân tử: Tập trung vào protein đích của các thuốc có chỉ định trên ĐTDĐ và một số enzym quan trọng trong bệnh sinh ĐTDĐ. Các tiêu chí lựa chọn bao gồm: protein được xác định bằng phương pháp khúc xạ tia X, độ phân giải tinh chỉnh từ 1,5 - 4,0 Å, gen không đột biến, chỉ số RMSD giữa phối tử tự nhiên (native ligand) trước và sau khi dock nhỏ hơn 3. Các đích tác động sau khi được lựa chọn sẽ được tải về từ cơ sở dữ liệu PDB. Sử dụng phần mềm Discovery Studio Visualiser được dùng để tách riêng protein với ligand đồng kết tinh thành 2 tệp riêng biệt định dạng ".pdb". Phần mềm Autodock tool giúp tạo grid box và lưu thông số vào tệp "config.txt". Tệp protein sau đó được lưu dưới định dạng ".pdbqt". Cấu trúc của các ligand được tải về từ Pubchem dưới định dạng ".sdf". Sử

dùng hệ điều hành ubuntu chuyển toàn bộ tệp ligand từ định dạng “.sdf” sang “.pdbqt”. Quy trình thực hiện gắn kết phân tử với nhiều ligand bằng Autodock Vina cùng sự hỗ trợ của ubuntu. Kết quả sau khi gắn kết phân tử được đánh giá thông qua hai tiêu chí: Năng lượng gắn kết (kcal.mol<sup>-1</sup>) và tương tác giữa hợp chất và các acid amin của đích tác động trong khoang gắn kết.

## 2.7. Xử lý và phân tích số liệu

- Số liệu của nghiên cứu được xử lý bằng phần mềm Microsoft Excel 365, cấu trúc 2D của các hợp chất được vẽ bằng phần mềm ChemDraw Pro 12.0.

## 3. KẾT QUẢ NGHIÊN CỨU

### 3.1. Thu thập và xử lý dữ liệu

Dữ liệu đầu vào gồm 159 TD và 739 TÂ đạt các tiêu chí lựa chọn và tiêu chí loại trừ. Trong đó, TD bao gồm 57 loại thuốc đã được FDA chấp thuận, 16 thuốc pha IV, 28 thuốc pha III, 49 thuốc pha II, 8 thuốc pha I và 1 thuốc pha đầu I. Các pha lâm sàng của thuốc TD được phân chia vào các tập huấn luyện và tập kiểm tra với tỉ lệ 80:20 để đảm bảo sự cân bằng dữ liệu. Các thuốc TÂ sau khi được phân chia vào 5 tập dữ liệu con cũng được tiếp tục phân chia thành tập huấn luyện và tập kiểm tra với cùng tỉ lệ 80:20 theo ngẫu nhiên và được gộp cùng với thuốc TD thành 10 tập dữ liệu hoàn chỉnh (5 tập huấn luyện và 5 tập kiểm tra). Dữ liệu về hợp chất tự nhiên gồm 336 hợp chất có hoạt tính trên ĐTĐ từ cơ sở dữ liệu DiaNatDB làm dữ liệu kiểm tra, 13.105 hợp chất từ CMAUP và 1.000 hợp chất từ Vietherb làm đối tượng sàng lọc.

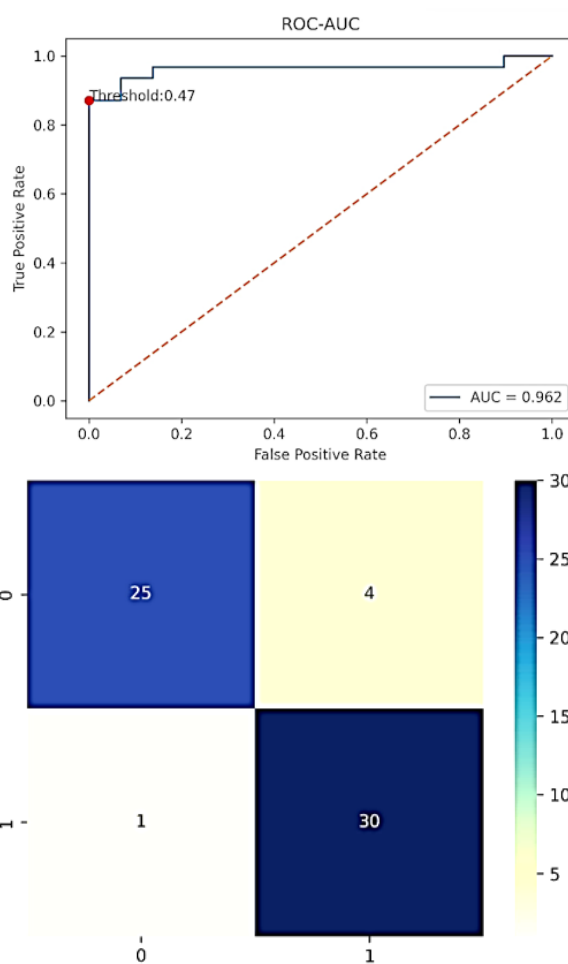
### 3.2. Chuyển đổi cấu trúc SMILES thành mã nhị phân

Sau khi phân chia dữ liệu đầu vào thành các tập huấn luyện và kiểm tra, cấu trúc canonical SMILES của thuốc được chuyển đổi thành mã nhị phân bằng tổ hợp 15 dấu vân tay phân tử. Cụ thể hơn, việc tạo các tổ hợp được dựa trên quy luật chọn 1 trong 4, chọn 2 trong 4, chọn 3 trong 4 và chọn 4 trong 4 dấu vân tay phân tử ban đầu mà không có tổ hợp nào trùng nhau về thành phần dấu vân tay phân tử. Số lượng đặc trưng của 4 dấu vân tay phân tử đơn lẻ ECFP4, FCFP4, MACCS và RDK lần lượt là 1024, 1024, 166 và 2048. Sau khi tổ hợp, số lượng đặc trưng của tổ hợp sẽ bằng tổng số lượng đặc trưng của các thành phần dấu vân tay có trong tổ hợp đó. Các đặc trưng quan trọng nhất được giữ lại sau trích chọn đặc trưng bằng thuật toán ImRMR được lựa chọn để xây dựng mô hình học máy bằng thuật toán XGBoost.

### 3.3. Xây dựng mô hình dự đoán các hợp chất tự nhiên có hoạt tính kháng đái tháo đường - mô hình DiMeNP

Sau khi trích chọn đặc trưng, tổ hợp dấu vân tay phân tử FM cho mô hình có hiệu suất tốt nhất với các giá trị độ nhạy là 0,968, độ đặc hiệu 0,862, độ chính xác 0,917,

độ tinh xác 0,882, điểm F1 0,923 và điểm MCC 0,837.



Hình 1. AUC-ROC và ma trận nhầm lẫn của mô hình DiMeNP

Giá trị AUC-ROC 0,962 và ma trận nhầm lẫn của mô hình từ tổ hợp FM (Hình 1) cho thấy chỉ có 1 trường hợp bị dự đoán là âm tính giả (thuốc có hoạt tính nhưng được dự đoán là không có). Do đó, mô hình FM được lựa chọn làm mô hình DiMeNP nhờ hạn chế được khả năng bỏ qua hợp chất tiềm năng trong thực tế. Mô hình DiMeNP dự đoán được 215 trong 336 hợp chất của cơ sở dữ liệu DiaNatDB có hoạt tính điều trị ĐTĐ. Mặc dù mô hình không dự đoán một số hợp chất là có hoạt tính, điều này không có nghĩa mô hình không hiệu quả. Một số lý do có thể kể đến bao gồm các hợp chất trên cơ sở dữ liệu ban đầu được thu thập với tiêu chí không tập trung vào hoạt tính trị ĐTĐ được sử dụng trong y học dân gian bất kể mục tiêu sinh học. Trong 121 hợp chất được dự đoán là không có hoạt tính, 27 hợp chất không có thông tin rõ ràng về hoạt tính trên ĐTĐ. Hơn thế nữa, các hợp chất chưa được thử nghiệm lâm sàng mà chỉ ở các thử nghiệm *in vitro* hoặc *in vivo*.

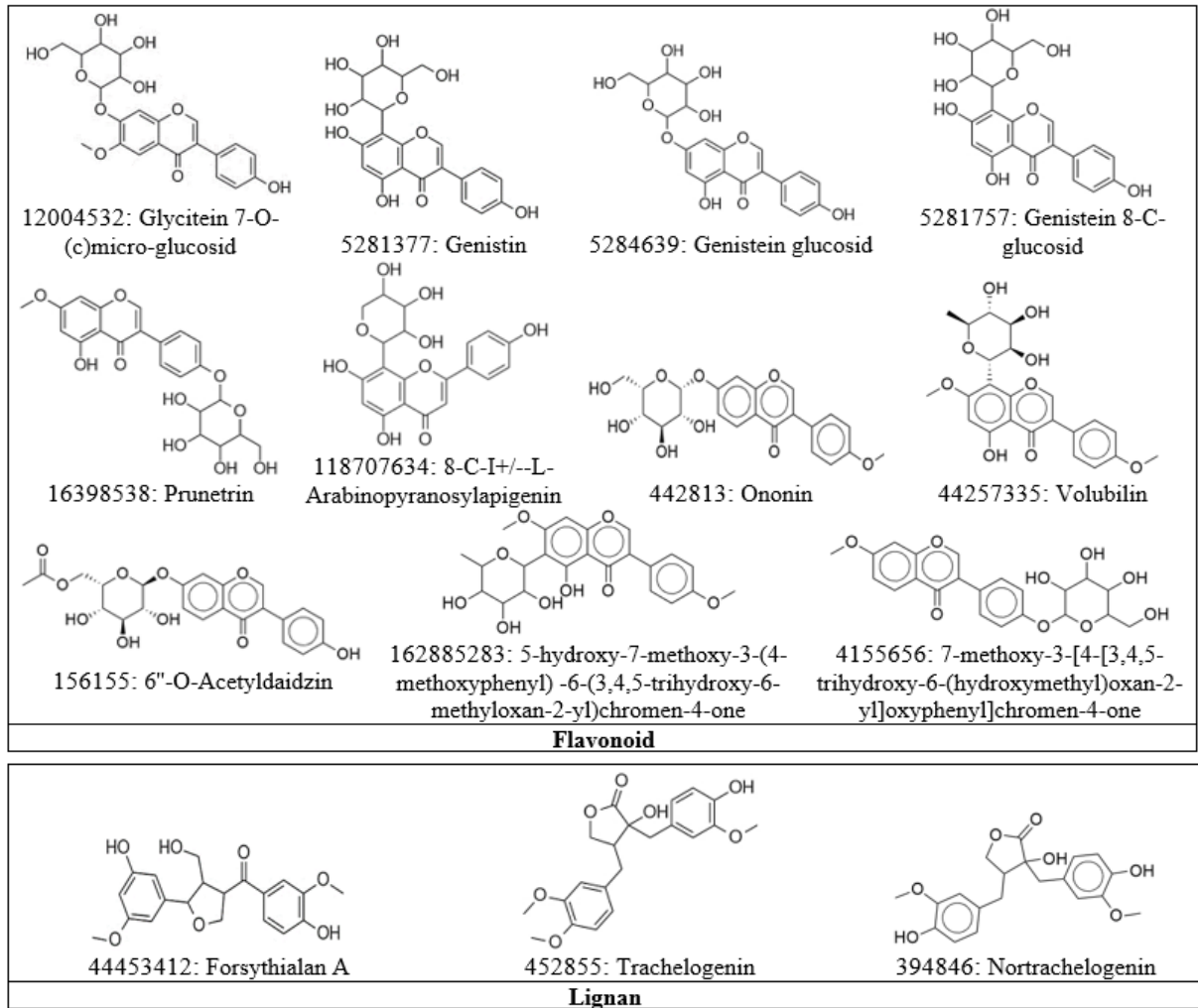
### 3.4. Sàng lọc hợp chất từ cơ sở dữ liệu CMAUP và Vietherb

Trên cơ sở dữ liệu CMAUP, mô hình sàng lọc được 49 hợp chất có tiềm năng điều trị ĐTĐ với giá trị xác suất



> 0,9 được dự đoán là có hoạt tính. Trong đó, 8 hợp chất đáp ứng tiêu chí về sinh khả dụng và tính giống thuốc được tiếp tục đánh giá bằng phương pháp gắn kết phân tử. Tương tự, từ cơ sở dữ liệu Vietherb, mô hình sàng lọc được 7 hợp chất thỏa các tiêu chí trên. Vì có một hợp chất bị trùng lặp giữa hai cơ sở dữ liệu,

tổng cộng 14 hợp chất tự nhiên được tiếp tục đánh giá bằng phương pháp gắn kết phân tử. Dựa trên khung cấu trúc hóa học chính, các chất này có thể được phân loại thành các nhóm hợp chất tự nhiên bao gồm flavonoid và lignan (Hình 2).



Hình 2. Cấu trúc 2D của các hợp chất được sàng lọc

### 3.5. Thu thập đích tác động và gắn kết phân tử

Các nhóm thuốc gây hạ glucose máu nặng bao gồm sulfonyleurea và meglitinid bị loại bỏ. Nhóm thuốc SGLT2 cũng không được lựa chọn do vi phạm tiêu chí lựa chọn với phương pháp thực nghiệm là phương pháp sử dụng kính hiển vi điện tử. Cuối cùng, 5 protein được lựa chọn cho quá trình gắn kết phân tử bao gồm PPAR- $\gamma$ ,  $\alpha$ -glucosidase,  $\alpha$ -amylase, GLP1 và DPP4. Chứng dương được sử dụng là thuốc trị ĐTĐ metformin. Kết quả gắn kết cho thấy cả 14 hợp chất đều liên kết tốt trên 5 đích tác động với năng lượng liên kết thuộc khoảng -12,9 đến -6,4 kcal.mol<sup>-1</sup> (Bảng 1). Năng lượng liên kết của metformin thuộc khoảng -5,3 đến -4,8 kcal.mol<sup>-1</sup>, cho thấy ái lực liên kết là thấp hơn đáng kể trên cả 5 đích tác động so với các hợp chất sàng lọc được.

Bảng 1. Kết quả gắn kết phân tử của 14 hợp chất trên 5 đích tác động tiềm năng

Protein/ PubChem ID	GLP-1	DPP	PPAR- $\gamma$	$\alpha$ -glucosidase	$\alpha$ -amylase
12004532	-9,5	-8,5	-6,6	-7,7	-9,4
5284639	-9,6	-8,6	-7,7	-10,2	-9,3
118707634	-9,2	-8,1	-7,6	-7,6	-8,7
44453412	-9,3	-7,9	-6,4	-8,5	-7,8

Protein/ PubChem ID	GLP-1	DPP	PPAR- $\gamma$	$\alpha$ -glucosidase	$\alpha$ -amylase
5281757	-9,6	-8,3	-7,8	-8,2	-9,4
5281377	-9,8	-8,8	-7,5	-9,9	-9,2
16398538	-9,4	-7,8	-6,9	-8,4	-9,1
156155	-10,0	-8,7	-7,6	-9,9	-9,6
44257335	-10,0	-8,1	-7,3	-7,9	-8,8
442813	-9,4	-8,0	-7,2	-7,6	-9,0
452855	-9,2	-7,9	-6,4	-8,5	-7,9
394846	-9,1	-7,3	-7,6	-9,0	-8,0
16288528	-12,6	-11,7	-12,0	-12,9	-11,4
4155656	-11,9	-10,9	-11,9	-12,3	-11,2
Metformin	-4,9	-5,3	-4,9	-5,1	-4,8

## 4. BÀN LUẬN

### 4.1. Bàn luận về kết quả sàng lọc bằng mô hình học máy và gắn kết phân tử

Các hợp chất được sàng lọc từ mô hình đều có khả năng gắn kết mạnh mẽ với các mục tiêu sinh học tiềm năng trong điều trị ĐTĐT2. Cụ thể, các hợp chất này có ái lực gắn kết tốt nhất trên GLP1 và vượt trội hơn chúng dương metformin. GLP1 có vai trò giúp tăng tiết insulin, ức chế tiết glucagon và cho các lợi ích trên biến chứng tim mạch. Ngoài ra đây còn là một trong nhiều đích tác động của metformin [6]. Điều này cũng củng cố tiềm năng tác động trên mục tiêu GLP1 của các hợp chất đã được mô hình DiMeNP dự đoán. Đồng thời nhấn mạnh tiềm năng của các hợp chất trong điều trị ĐTĐT2 thông qua nhiều con đường sinh hóa khi có khả năng gắn kết trên đa mục tiêu có lợi và giảm thiểu các biến chứng liên quan.

### 4.2. Ứng dụng và hạn chế của nghiên cứu

Kết quả sàng lọc của mô hình có thể trở thành nền tảng cho các nghiên cứu *in vitro*, *in vivo* và lâm sàng. Với các kết quả đạt được, DiMeNP có thể mở ra cơ hội mới cho nghiên cứu và phát triển các thuốc điều trị đái tháo đường từ dược liệu, mang lại giá trị tiền đề cho một phương pháp điều trị hiệu quả và an toàn hơn. Để tối ưu hóa mô hình, cần có những nghiên cứu toàn diện hơn trong dữ liệu đầu vào về đích tác động tiềm năng mới trong bệnh sinh ĐTĐT; đích tác động của các thuốc và hợp chất; mối liên hệ giữa thành phần, hàm lượng của

hợp chất và dược liệu chứa chúng. Qua đó, mô hình có thể cung cấp tiền đề sinh học phân tử cho các nghiên cứu phát triển phương pháp điều trị ĐTĐT từ việc phối hợp dược liệu có tương tác hiệp đồng ở các đích tác động khác nhau trên con đường bệnh sinh ĐTĐT.

## 5. KẾT LUẬN

Mô hình DiMeNP chứng minh được độ tin cậy với độ chính xác là 0,917 và AUC là 0,962 cùng với khả năng dự đoán được 215 trên tổng 336 hợp chất có hoạt tính trong tập dữ liệu kiểm tra. Từ 2 cơ sở dữ liệu với 13.105 hợp chất ban đầu, mô hình đã sàng lọc được 14 hợp chất tự nhiên tiềm năng với tác động đa mục tiêu trong điều trị đái tháo đường thuộc 2 nhóm cấu trúc flavonoid và lignan, có xác suất dự đoán > 0,9 và thỏa các tiêu chí về sinh khả dụng cũng như tính giống thuốc. Tất cả 14 hợp chất có khả năng gắn kết tốt trên các đích tác động tiềm năng. Bằng cách áp dụng các công cụ mạnh mẽ như học máy và gắn kết phân tử, nghiên cứu đã đưa ra một cách tiếp cận mới, tiết kiệm chi phí và nhanh chóng để sàng lọc và dự đoán hoạt tính của các hợp chất tự nhiên trong điều trị ĐTĐT.

## TÀI LIỆU THAM KHẢO

- [1] Facts & Figures Available online: <https://idf.org/about-diabetes/diabetes-facts-figures/> (accessed on 6 December 2023).
- [2] Dhankhar, S.; Chauhan, S.; Mehta, D.K.; Nitika; Saini, K.; Saini, M.; Das, R.; Gupta, S.; Gautam, V. Novel Targets for Potential Therapeutic Use in Diabetes Mellitus. *Diabetol. Metab. Syndr.* 2023, 15, 17, doi:10.1186/s13098-023-00983-5.
- [3] Dowarah, J.; Singh, V.P. Anti-Diabetic Drugs Recent Approaches and Advancements. *Bioorg. Med. Chem.* 2020, 28, 115263, doi:10.1016/j.bmc.2019.115263.
- [4] O'Connell, J.M.; Manson, S.M. Understanding the Economic Costs of Diabetes and Prediabetes and What We May Learn About Reducing the Health and Economic Burden of These Conditions. *Diabetes Care* 2019, 42, 1609–1611, doi:10.2337/dci19-0017.
- [5] Makhoba, X.H.; Viegas Jr, C.; Mosa, R.A.; Viegas, F.P.D.; Pooe, O.J. Potential Impact of the Multi-Target Drug Approach in the Treatment of Some Complex Diseases. *Drug Des. Devel. Ther.* 2020, 14, 3235–3249, doi:10.2147/DDDT.S257494.
- [6] Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. GeneCards: a novel functional genomics compendium with automated data mining and query reformulation support. *Bioinformatics.* 1998;14(8):656-64. doi: 10.1093/bioinformatics/14.8.656. PMID: 9789091.