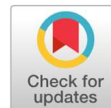




ISSN : 1859-1779

Nghiên cứu Dược học

Tạp chí Y học Thành phố Hồ Chí Minh - Dược học;29(2):11-19
<https://doi.org/10.32895/hcjm.p.2026.02.02>



Sàng lọc chất chủ vận thụ thể apelin dựa trên độ tương đồng phân tử và tối ưu hóa hiệu năng thông qua kỹ thuật voting-stacking

Tô Văn Thịnh¹, Nguyễn Song Như Ngọc¹, Nguyễn Văn Phước Chung¹, Vũ Hải Đăng¹, Trương Ngọc Tuyền¹, Nguyễn Quý Hiến^{1,*}

¹Trường Dược, Đại học Y Dược Thành phố Hồ Chí Minh, Thành phố Hồ Chí Minh, Việt Nam

Tóm tắt

Đặt vấn đề: Apelin, một peptid mang hoạt tính sinh học, là phối tử nội sinh của thụ thể Apelin (APJ) liên kết protein G (APLNR) tìm thấy chủ yếu ở hệ thần kinh trung ương và các mô ngoại biên. Chức năng và tác dụng điều trị các bệnh mạn tính không lây nhiễm của APJ đã được công bố rất nhiều trong hai thập kỷ vừa qua. Nghiên cứu đánh giá khả năng ứng dụng của độ tương đồng phân tử trong sàng lọc ảo các chất chủ vận thụ thể APJ.

Mục tiêu: Sàng lọc các chất chủ vận thụ thể Apelin bằng cách ứng dụng độ tương đồng phân tử dựa trên các dấu vân tay phân tử, đồng thời tối ưu hóa hiệu năng sàng lọc thông qua các phương pháp học tập kết hợp Voting và Stacking.

Đối tượng và phương pháp nghiên cứu: Ba chất tham chiếu có hoạt tính chủ vận thụ thể Apelin mạnh (CMF-019, AMG-986, BMS-986224) được dùng để đánh giá khả năng nhận diện chất có hoạt tính dựa trên độ tương đồng phân tử thông qua tính toán 12 loại dấu vân tay và hệ số tương đồng Tanimoto. Phương pháp học tập kết hợp Voting và Stacking cũng được áp dụng nhằm cải thiện hiệu năng tổng thể. Khả năng sàng lọc chất chủ vận thụ thể Apelin dựa trên độ tương đồng được đánh giá dựa trên các chỉ số AUC, EF, GH và F1.

Kết quả: Phương pháp dựa trên độ tương đồng phân tử cho thấy hiệu năng sàng lọc cao; việc tích hợp phương pháp học tập kết hợp cải thiện hiệu năng sàng lọc hướng chủ vận thụ thể Apelin, với kết quả đánh giá ngoại AUC = 0,997; GH= 0,924 ; F1= 0,875

Từ khóa: Apelin/APJ; bệnh mạn tính không lây nhiễm; sàng lọc ảo; độ tương đồng phân tử; Voting; Stacking

Abstract

MOLECULAR SIMILARITY FOR SCREENING APELIN AGONISTS: EVALUATION AND PERFORMANCE OPTIMIZATION WITH VOTING-STACKING TECHNIQUES

To Van Thinh, Nguyen Song Nhu Ngoc, Nguyen Van Phuoc Chung, Vu Hai Dang, Truong Ngoc Tuyen, Nguyen Qui Hien

Background: Apelin, a bioactive peptide, is the endogenous ligand of the G protein-coupled receptor APJ (APLNR) found mostly in the central

Ngày nhận bài: 08-10-2025 / Ngày chấp nhận đăng bài: 12-01-2026 / Ngày đăng bài: 28-02-2026

*Tác giả liên hệ: Nguyễn Quý Hiến. Trường Dược, Đại học Y Dược Thành phố Hồ Chí Minh, Thành phố Hồ Chí Minh, Việt Nam.

E-mail: nguyequihien@ump.edu.vn

© 2026 Bản quyền thuộc về Tạp chí Y học Thành phố Hồ Chí Minh.

nervous system (CNS) and other peripheral tissues. Studies on biological effects and therapeutic values towards chronic non-communicable diseases have been widely published throughout the last two decades. This study aims to compute and evaluate the applicability of Molecular Similarity for performing virtual screening (VS) of APJ agonists.

Objective: To screen Apelin receptor agonists by applying molecular similarity based on molecular fingerprints, while optimizing screening performance through ensemble learning methods, including Voting and Stacking.

Methods: Three standard reference compounds with strong Apelin agonist activity (CMF-019, AMG-986, BMS-986224) were used to evaluate similarity-based ability to identify active compounds by computing 12 types of fingerprints and the Tanimoto coefficient. The ensemble learning methods including Voting and Stacking were also applied to improve the overall performance. The similarity-based screening performance for Apelin receptor agonists was evaluated using the AUC, EF, GH, and F1 metrics.

Results: The molecular similarity approach delivers strong screening performance; incorporating ensemble learning methods further enhances this capability, yielding external validation metrics of AUC = 0.997, GH = 0.924, and F1 = 0.875.

Keywords: Apelin/APJ; chronic non-communicable diseases; virtual screening; molecular similarity; Voting; Stacking

1. ĐẶT VẤN ĐỀ

Sự phát triển nhanh chóng của xã hội cùng với những biến đổi lớn về môi trường sống và khí hậu đã khiến các bệnh mạn tính không lây nhiễm ngày càng khởi phát sớm và phổ biến trên toàn cầu [1]. Thời gian qua, có rất nhiều nhóm thuốc mới ra đời với mục tiêu điều trị hay cải thiện sức khỏe, chất lượng sống cho người bệnh, tuy nhiên, tình trạng không đáp ứng thuốc xuất hiện ngày càng nhiều đặt ra yêu cầu phát triển các hướng điều trị mới nhằm đảm bảo cho quá trình điều trị đạt hiệu quả, đồng thời giảm thiểu tối đa nguy cơ các biến chứng. Nhờ các kỹ thuật thiết kế thuốc có sự trợ giúp của máy tính (Computer-Aided Drug Design - CADD) mà các nhà khoa học đã giảm đáng kể được thời gian và chi phí để đưa thuốc mới ra thị trường. Trong đó, quá trình sàng lọc ảo (Virtual Screening - VS) là một trong những phương pháp CADD thực hiện ở các bước đầu trong quá trình khám phá thuốc để chọn lựa được các hợp chất có tiềm năng tương tác với đích tác động [2]. Đối với sàng lọc ảo dựa trên phối tử, độ tương đồng phân tử thường được tính toán dựa trên dấu vân tay phân tử và so sánh bằng hệ số Tanimoto, cho phép xếp hạng nhanh các hợp chất tiềm năng trong những cơ sở dữ liệu lớn [3]. Vì mỗi loại dấu vân tay phản ánh các khía cạnh khác nhau của cấu trúc nên các chiến lược học tập kết hợp có thể được phối hợp để khai thác tính bổ sung giữa các dấu vân tay khác nhau. Các kỹ thuật như Voting có thể thực hiện hợp nhất điểm và xếp hạng từ nhiều phép tìm kiếm tương đồng để tăng độ ổn định và hiệu năng, trong khi Stacking dùng mô hình meta học từ đầu ra của nhiều mô hình nền để tăng khả năng khái quát [4,5].

Apelin, một peptid có hoạt tính sinh học, là phối tử nội sinh của thụ thể APJ thuộc nhóm thụ thể kết hợp protein G, được biểu hiện rộng rãi các mô ngoại vi và trong hệ thống thần kinh trung ương. Hệ thống Apelin/APJ giữ vai trò điều hòa nhiều chức năng sinh lý và là đích điều trị tiềm năng của các bệnh mạn tính không lây nhiễm [6-8]. Apelin tăng cường co bóp cơ tim, giảm xơ hóa và được xem như một dấu ấn sinh học tiềm năng; đồng thời cải thiện sử dụng glucose, tăng nhạy insulin và tham gia điều hòa bệnh đái tháo đường; ngoài ra còn bảo vệ tế bào thần kinh trong các bệnh thoái hóa như Alzheimer, Parkinson và Huntington [9-13]. Từ thực tiễn đó, nghiên cứu này tập trung xây dựng phương pháp tương đồng phân tử ứng dụng trong sàng lọc ảo nhằm tìm kiếm các chất tiềm năng hướng đồng vận APJ.

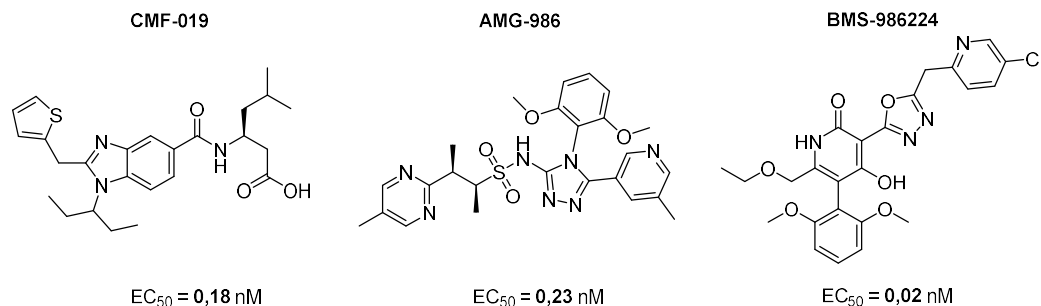
APJ là một đích sinh học quan trọng liên quan đến nhiều bệnh lý phức tạp, nhưng các nghiên cứu thiết kế thuốc nhắm mục tiêu này vẫn còn hạn chế. Nghiên cứu của Suvankar Banerjee (2022) đã ứng dụng QSAR (Quantitative Structure-Activity Relationship - Tương quan cấu trúc - tác dụng) và docking để xác định các nhóm thế quan trọng (điển hình như nhóm 4-chlorobenzyl, 4-bromobenzyl và 2-methoxy) và các acid amin chủ chốt tham gia tương tác với APJ [14]. Trên cơ sở này, Tran và cộng sự (2024) phát triển mô hình pharmacophore 3D kết hợp phân cụm Butina và học tập kết hợp (ensemble learning), qua đó mở rộng không gian dữ liệu, tối ưu hóa cấu trúc mô hình và đạt hiệu suất sàng lọc cao [15]. Những kết quả này củng cố vai trò ngày càng rõ nét của các phương pháp tính toán trong nghiên cứu và thiết kế thuốc nhắm trúng đích APJ.

2. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Đối tượng nghiên cứu

Nghiên cứu được tiến hành trên cơ sở dữ liệu tổng hợp được từ các bài báo khoa học và các bằng sáng chế (WO2016187308A1, WO2018093576A1, WO2019169193A1, WO2019036024A1, WO2016176476A1,

WO2017096130A1, WO2018071622A1, US9156796B2) thu được tập hoạt tính bao gồm 6944 chất thỏa mãn đồng thời 3 tiêu chí: đích tác động là thụ thể APJ ở người, có hoạt tính chủ vận với APJ, và có $EC_{50} < 100$ nM. Nghiên cứu sử dụng ba chất chủ vận thụ thể Apelin phân tử nhỏ là CMF-019, AMG-986 (AZELAPRAG) và BMS-986224 làm chất tham chiếu với giá trị EC_{50} lần lượt là 0,18 nM, 0,23 nM và 0,02 nM [16-18] (Hình 1).



Hình 1. Ba chất tham chiếu xây dựng phương pháp phân loại dựa trên độ tương đồng

2.2. Phương pháp nghiên cứu

2.2.1. Tài nguyên nghiên cứu

Nghiên cứu *in silico* được tiến hành tại Bộ môn Hóa Hữu cơ, Trường Dược, Đại học Y Dược Thành phố Hồ Chí Minh trên máy tính sử dụng hệ điều hành Linux 22.04, CPU AMD Ryzen 9 3900X, 12 nhân, tốc độ xử lý 3,79 Hz, bộ nhớ trong 512 GB, RAM 96,0 GB với card đồ họa AMD Radeon RX 6900 XT. Ngôn ngữ lập trình Python được sử dụng để tính toán và xử lý kết quả.

2.2.2. Chuẩn bị dữ liệu

Tập hoạt tính sau đó trải qua các bước: (1) chuẩn hóa SMILES thành Canonical SMILES, (2) biến đổi EC_{50} thành pEC_{50} và (3) loại bỏ các chất trùng lặp. Để đảm bảo tính giống thuốc, quy tắc 5 Lipinski (RO5) được áp dụng để sàng lọc các tác nhân điều trị bằng đường uống và chọn cấu trúc đáp ứng ít nhất ba quy tắc (vi phạm một quy tắc). Tập hoạt tính được đưa vào DeepCoy để tạo sinh các “mồi nhử” với tỷ lệ 1:50, từ đó kết hợp cùng dữ liệu gốc tạo thành một tập đánh giá hoàn chỉnh.

2.2.3. Tính toán và đánh giá khả năng ứng dụng của độ tương đồng phân tử

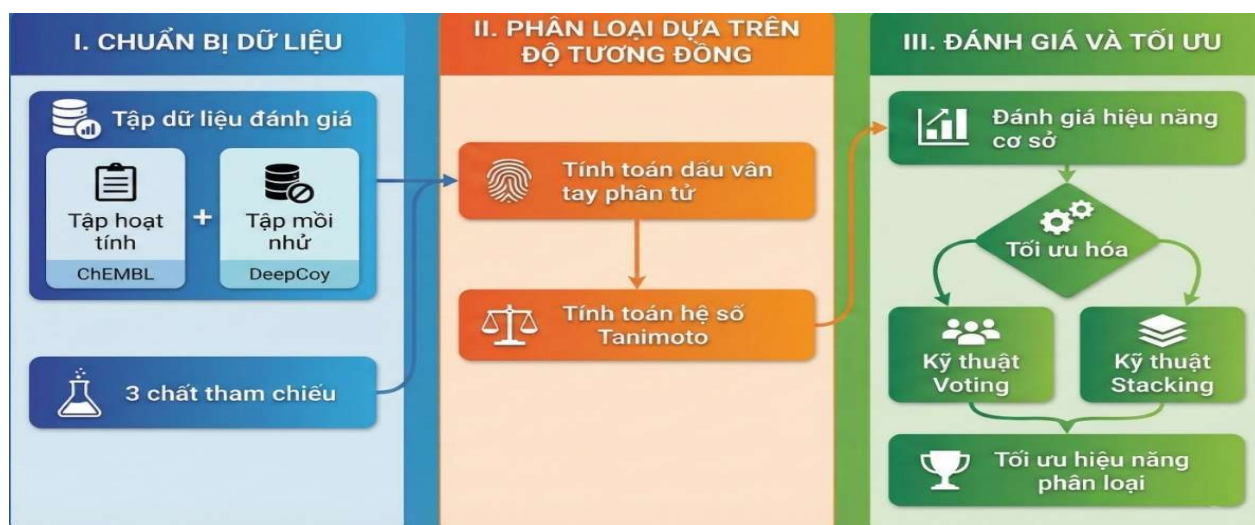
Tính toán độ tương đồng phân tử

Quá trình xử lý dữ liệu, tính toán dấu vân tay độ tương đồng phân tử được minh họa ở Hình 2.

Nghiên cứu tiến hành xây dựng bộ dữ liệu hệ số tương đồng Tanimoto của 36 cặp dấu vân tay - chất tham chiếu, tương ứng với 12 loại dấu vân tay và 3 chất tham chiếu CMF-019, AMG-986, BMS-986224. Đầu tiên, nghiên cứu thực hiện tính toán 12 dấu vân tay phân tử bao gồm MACCS, AVALON, MAP4, RDK5, RDK6, RDK7, ECFP2, ECFP4, ECFP6, FCFP2, FCFP4 và FCFP6 của ba chất tham chiếu và toàn bộ tập đánh giá. Sau đó, nhóm nghiên cứu tiếp tục tính toán hệ số tương đồng Tanimoto giữa chất tham chiếu và dữ liệu tập đánh giá.

Đánh giá hiệu năng phân loại

Tập đánh giá được sử dụng để đánh giá hiệu năng phân biệt các chất hoạt tính và mồi nhử thông qua dấu vân tay và phương pháp tương đồng. Hai thông số chính được sử dụng là hệ số làm giàu tại 1% (EF1%) [19] và diện tích dưới đường cong ROC (AUC-ROC) [20]. Trong đó, EF1% phản ánh mức độ nhận diện sớm các chất hoạt tính trong 1% số chất có thứ hạng cao nhất của kết quả sàng lọc. Khả năng phân biệt được xem là đạt yêu cầu khi đồng thời thỏa mãn $AUC-ROC > 0,7$ và $EF1\% > 1$ [19,20]. Ngoài ra, nghiên cứu còn sử dụng thêm điểm số Güner-Henry (GH) và F1-Score nhằm đánh giá toàn diện hiệu năng phân loại, với ngưỡng chấp nhận lần lượt là $GH > 0,5$ và $F1 > 0,5$ [21,22].



Hình 2. Quy trình đánh giá khả năng sàng lọc chất chủ vận thụ thể Apelin dựa trên độ tương đồng

2.2.4. Tối ưu hóa quá trình ứng dụng độ tương đồng phân tử

Nghiên cứu xác định cặp chất tham chiếu-dấu vân tay tối ưu (cặp đối chứng nền) bằng cách chọn tổ hợp mang lại giá trị AUC-ROC cao nhất và vượt trội có ý nghĩa thống kê so với các lựa chọn còn lại. Để đảm bảo tính khách quan, tập đánh giá được chia 80:20 thành đánh giá nội và đánh giá ngoại. Đánh giá nội được biến thể thành đánh giá chéo 10 lần lặp lại 3 lần (10 x 3), sau đó kết quả được so sánh bằng kiểm định Wilcoxon. Đánh giá ngoại đối chiếu hiệu năng giữa cặp đối chứng nền và kết quả sau khi tối ưu hóa.

Hai kỹ thuật tối ưu hóa được áp dụng là hai phương pháp học tập kết hợp Voting và Stacking. Kỹ thuật Voting lấy ý tưởng từ phương pháp CHA (Common Hits Approach) và thuật toán Ensemble, với bản chất của phương pháp là tổ hợp sự có mặt một hay nhiều bit của một loại dấu vân tay nhất định trên ba chất tham chiếu (với trọng số của ba chất này là như nhau). Đầu vào kỹ thuật này yêu cầu một bộ dấu vân tay “phù hợp” nhất với cả ba chất tham chiếu, xác định thông qua kết quả AUC-ROC trung bình. Quá trình ứng dụng Voting vào tối ưu hóa trải qua ba bước bao gồm việc lựa chọn và tính toán dấu vân tay cho ba chất tham chiếu từ đó suy ra CHA của từng bit. Tiếp theo, nghiên cứu tiến hành tính toán hệ số tương đồng giữa CHA và tập đánh giá, và cuối cùng đánh giá khả năng phân loại thông qua đánh giá chéo trên tập đánh giá nội.

Kỹ thuật Stacking tối ưu hóa quá trình ứng dụng độ tương đồng phân tử bao gồm ba bước chính. Trước hết, một bộ dữ liệu mới được xây dựng dựa trên đầu ra của các phương pháp tính độ tương đồng sẵn có; trong trường hợp Optimal

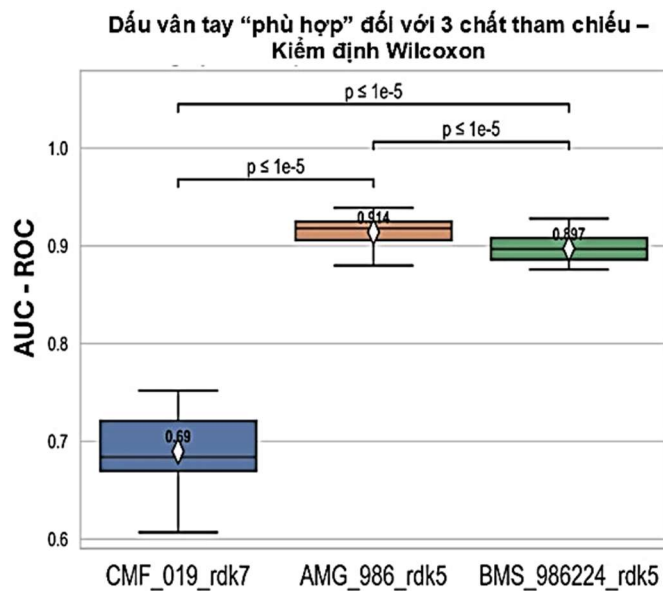
Stacking, thuật toán Random Forest [20,21] được áp dụng để thực hiện quá trình lựa chọn đặc trưng. Tiếp theo, meta-model được huấn luyện trên bộ dữ liệu mới nhằm học cách kết hợp thông tin từ các mô hình cơ sở. Cuối cùng, hiệu quả của quá trình tối ưu hóa được đánh giá thông qua đánh giá chéo trên tập đánh giá nội và kết quả được so sánh bằng kiểm định Wilcoxon.

3. KẾT QUẢ

3.1. Tính toán, đánh giá và lựa chọn cặp đối chứng nền

Nghiên cứu thu được 36 bộ dữ liệu độ tương đồng phân tử giữa tập đánh giá và các chất tham chiếu. Nghiên cứu tiến hành đánh giá chéo và thực hiện kiểm định Wilcoxon để xác định dấu vân tay “phù hợp” nhất cho từng chất tham chiếu.

Sau quá trình đánh giá chéo nội 10 lần lặp lại 3 lần, giá trị diện tích dưới đường cong AUC-ROC cao nhất khác biệt có ý nghĩa thống kê ở từng bộ dữ liệu lần lượt là: RDK7 (CMF-019) với $AUC-ROC = 0,690 \pm 0,036$; RDK5 (AMG-986) với $AUC-ROC = 0,914 \pm 0,015$; RDK5 (BMS_986224) với $AUC-ROC = 0,897 \pm 0,013$. Khi so sánh 3 bộ dữ liệu tương ứng với dấu vân tay tốt nhất, kết quả cho thấy bộ dữ liệu AMG-986 cho giá trị AUC-ROC cao nhất có ý nghĩa thống kê ($p < 0,05$) so với 2 bộ dữ liệu còn lại (Hình 3). Tổ hợp dấu vân tay RDK5 trên bộ dữ liệu AMG-986 được chọn làm cặp đối chứng nền với $AUC-ROC = 0,914 \pm 0,015$; $EF1\% = 50,070 \pm 0,215$; $GH = 0,803 \pm 0,01$; $F1 = 0,347 \pm 0,056$.

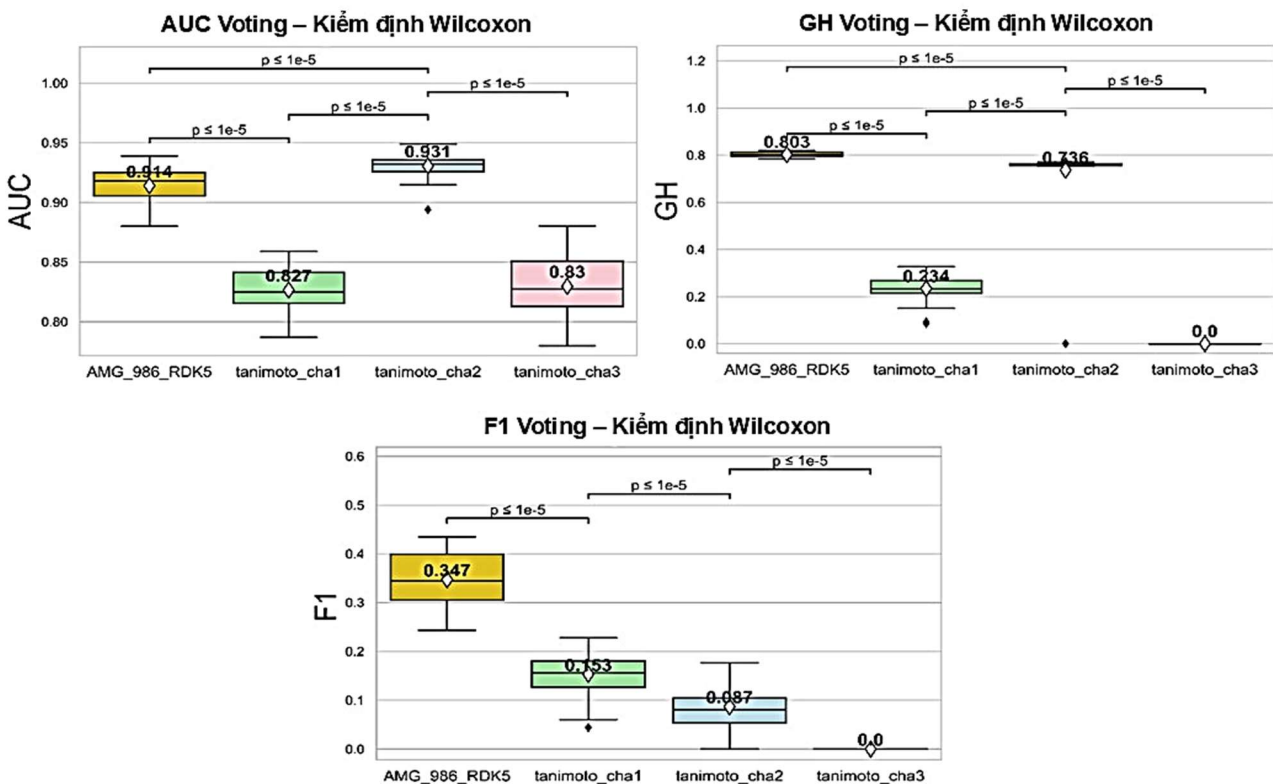


Hình 3. Biểu đồ hộp và râu mô tả 3 bộ dữ liệu bằng phương pháp đánh giá chéo sử dụng kiểm định Wilcoxon để so sánh dựa trên đại lượng AUC-ROC

3.2. Tối ưu hoá với phương pháp Voting

Giá trị AUC-ROC trung bình của 12 bộ dấu vân tay đối với ba chất tham chiếu. Kết quả cho thấy dấu vân tay RDK7 là dấu vân tay “phù hợp” nhất với cả ba bộ dữ liệu chất tham chiếu với hệ số tương đồng Tanimoto (AUC-ROC = 0,785).

Tiếp đó, nghiên cứu tiến hành tối ưu hóa quá trình ứng dụng độ tương đồng phân tử bằng kỹ thuật Voting. Mặc dù phương pháp CHA2 đạt giá trị dự đoán tổng thể cao nhất và có điểm AUC-ROC trung bình vượt trội so với cặp đối chứng nền (AMG-986_RDK5), nhưng khi xét đến hai chỉ số F1 và GH thì CHA2 vẫn cho kết quả thấp hơn (Hình 4).



Hình 4. Biểu đồ hộp và râu mô tả phương pháp Voting trên 3 đại lượng AUC-ROC, GH, F1

3.3. Tối ưu hoá với phương pháp Stacking

Nghiên cứu ứng dụng phương pháp Stacking nhằm tối ưu hóa khả năng phân loại dựa trên bộ dữ liệu chất tham chiếu AMG-986. Kết quả cho thấy mức độ cải thiện rõ rệt so với cặp đối chứng nền: điểm AUC-ROC trung bình tăng từ $0,914 \pm 0,015$ lên $0,996 \pm 0,004$, điểm GH từ $0,803 \pm 0,010$ lên $0,915 \pm 0,022$ và điểm F1 từ $0,347 \pm 0,056$ lên $0,864 \pm 0,034$ (Bảng 1). Các kết quả này cho kết quả có ý nghĩa thống kê với phép kiểm Wilcoxon ($p \leq 0,05$).

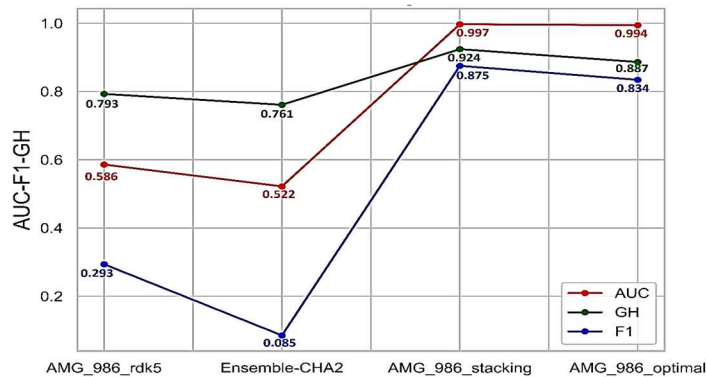
Bảng 1. Kết quả phương pháp Stacking trên bộ dữ liệu AMG-986.

Phương pháp Đại lượng		Phương pháp Stacking	Phương pháp chọn lọc đặc trưng bằng Random Forest	p-value
Đánh giá chéo nội	AUC-ROC	$0,996 \pm 0,004$	$0,992 \pm 0,005$	$p \leq 0,05$
	GH	$0,915 \pm 0,022$	$0,877 \pm 0,031$	
	F1	$0,864 \pm 0,034$	$0,825 \pm 0,036$	
Đánh giá ngoại	AUC-ROC	0,997	0,994	
	GH	0,924	0,887	
	F1	0,875	0,834	

Nghiên cứu sử dụng thuật toán Random Forest để lựa chọn các “đầu vôn tay quan trọng”. Random Forest đã chọn 7 đầu vôn tay với mức độ quan trọng giảm dần lần lượt là ECFP6, RDK5, Avalon, RDK7, MACCs, RDK6 và ECFP4 để tối ưu hóa. Kết quả cho thấy AUC-ROC, GH và điểm F1 của việc chọn lọc đặc trưng làm giảm (có ý nghĩa thống kê) hiệu năng phân loại dựa trên độ tương đồng phân tử so với khi chưa chọn lọc đặc trưng (Bảng 1). Có thể đưa ra kết luận rằng việc dùng chọn lọc đặc trưng không hiệu quả và có thể dẫn đến làm giảm hiệu năng phân loại.

3.4. Đánh giá ngoại

Để chứng minh cho tính tổng quát hoá, đánh giá ngoại được thực hiện ở giai đoạn cuối cùng để đánh giá khả năng phân loại bằng độ tương đồng phân tử một cách khách quan nhất. Kết quả đã chứng minh phương pháp Stacking không những giúp cải thiện kết quả có ý nghĩa thống kê trong đánh giá chéo, mà còn cải thiện hiệu năng trong đánh giá ngoại (Hình 5).



Hình 5. So sánh đại lượng AUC-ROC, GH, F1 giữa 4 giá trị đánh giá hiệu năng trên tập đánh giá ngoại

4. BÀN LUẬN

Kết quả đánh giá cho thấy hiệu năng phân loại sử dụng bộ đầu vôn tay của RDKit vượt trội hơn so với các đầu vôn tay khác ở cả ba chất tham chiếu. Khi quan sát ba kết quả tốt nhất tương ứng với từng chất tham chiếu, tổ hợp giữa chất tham chiếu AMG-986 với đầu vôn tay là RDK5 đạt hiệu năng tốt hơn có ý nghĩa thống kê và được chọn làm cặp chất đối chiếu nền. Nhằm đánh giá hiệu quả quá trình tối ưu hóa, nghiên cứu tiến hành so sánh các kết quả sau tối ưu (thu được từ các meta-model) với cặp chất đối chiếu nền. Kết quả cho

thấy kỹ thuật Voting làm giảm hiệu năng, trong khi đó kỹ thuật Stacking cho thấy khả năng cải thiện khả năng phân loại rõ rệt. Sự kết hợp giữa độ tương đồng phân tử và kỹ thuật Stacking không chỉ thể hiện hiệu năng tốt ở tập đánh giá nội mà còn trên tập đánh giá ngoại.

Sự thành công của phương pháp Stacking càng được củng cố khi so sánh với nghiên cứu của Tran và cộng sự (2024) [15]. Sự tương đồng về hiệu quả tối ưu của phương pháp Stacking trong hai hướng tiếp cận khác biệt - một bên dựa trên đặc trưng hóa học 2D và một bên dựa trên cấu trúc 3D - đã chứng minh cho tính ổn định và khả năng tổng quát

hóa của kỹ thuật Stacking. Do đó, việc tổng hợp đa chiều thông tin thông qua mô hình meta là chìa khóa để nâng cao hiệu năng khi so sánh với các phương pháp đơn lẻ, giúp nhận diện các chất chủ vận tiềm năng trong tập dữ liệu phức tạp. Điều này giúp khả năng ứng dụng của độ tương đồng phân tử trong phân biệt giữa các chất có hoạt tính và các môi như có cấu trúc tương tự trở nên tốt hơn, khắc phục được những điểm yếu cố hữu của từng cặp chất tham chiếu - dấu vân tay riêng lẻ.

5. KẾT LUẬN

Nghiên cứu đã đánh giá được khả năng phân biệt chất chủ vận thụ thể Apelin dựa trên độ tương đồng phân tử và tối ưu hóa bằng phương pháp học tập kết hợp Voting và Stacking. Kết quả đã chỉ ra rằng việc lựa chọn chất tham chiếu AMG-986 và loại dấu vân tay RDKit có ảnh hưởng quyết định đến hiệu năng phân loại. Đồng thời, nghiên cứu còn cho thấy kỹ thuật Stacking giúp tối ưu hóa hiệu năng phân loại tốt hơn khi so với các phương pháp riêng lẻ.

Trên cơ sở những kết quả đạt được, để hoàn thiện và phát triển nghiên cứu trong tương lai, chúng tôi sẽ tích hợp dấu vân tay mang thông tin cấu trúc 3D của các hợp chất vào tính toán độ tương đồng. Trong đó, phương pháp này sẽ được xây dựng dựa trên bộ dữ liệu với chất tham chiếu là AMG-986 bằng kỹ thuật stacking và thực hiện đánh giá, so sánh với phương pháp hiện tại. Đồng thời, mô hình docking phân tử trên thụ thể APJ sẽ được tiến hành song song với phương pháp tính toán độ tương đồng. Các phương pháp và mô hình sau khi hoàn thiện sẽ được ứng dụng trong sàng lọc ảo nhằm tìm kiếm các chất chủ vận tiềm năng trên thụ thể Apelin; các cấu trúc có triển vọng cao sẽ được lựa chọn để tiến hành tổng hợp và đánh giá hoạt tính sinh học trong các bước tiếp theo.

Nguồn tài trợ

Nghiên cứu này được tài trợ kinh phí bởi Đại học Y Dược Thành phố Hồ Chí Minh theo Hợp đồng số 218/2025/HĐ-ĐHYD, ngày 28 tháng 04 năm 2025.

Xung đột lợi ích

Không có xung đột lợi ích nào liên quan đến nghiên cứu này.

ORCID

Tô Văn Thịnh

<https://orcid.org/0000-0002-7640-0807>

Nguyễn Song Như Ngọc

<https://orcid.org/0009-0009-9764-9677>

Nguyễn Văn Phước Chung

<https://orcid.org/0009-0000-1825-2781>

Vũ Hải Đăng

<https://orcid.org/0009-0002-4023-6854>

Trương Ngọc Tuyền

<https://orcid.org/0000-0002-0952-1633>

Nguyễn Quý Hiền

<https://orcid.org/0000-0003-4907-2614>

Đóng góp của các tác giả

Ý tưởng nghiên cứu: Tô Văn Thịnh, Nguyễn Song Như Ngọc.

Đề cương và phương pháp nghiên cứu: Tô Văn Thịnh, Nguyễn Quý Hiền.

Thu thập dữ liệu: Nguyễn Văn Phước Chung, Vũ Hải Đăng.

Giám sát nghiên cứu: Nguyễn Quý Hiền, Trương Ngọc Tuyền.

Nhập dữ liệu: Nguyễn Văn Phước Chung, Vũ Hải Đăng.

Quản lý dữ liệu: Nguyễn Song Như Ngọc, Nguyễn Văn Phước Chung.

Phân tích dữ liệu: Tô Văn Thịnh, Nguyễn Song Như Ngọc.

Viết bản thảo đầu tiên: Tô Văn Thịnh, Nguyễn Song Như Ngọc.

Góp ý bản thảo và đồng ý cho đăng bài: Nguyễn Quý Hiền, Trương Ngọc Tuyền.

Cung cấp dữ liệu và thông tin nghiên cứu

Tác giả liên hệ sẽ cung cấp dữ liệu nếu có yêu cầu từ Ban biên tập.

Chấp thuận của Hội đồng Đạo đức

Nghiên cứu này miễn trừ Hội đồng Đạo đức.

TÀI LIỆU THAM KHẢO

1. Non communicable diseases. [cited 2025 Sep 25]. <https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases>.

2. Walters WP, Stahl MT, Murcko MA. Virtual screening-an overview. *Drug Discovery Today*. 1998;3(4):160-178. Doi:10.1016/S1359-6446(97)01163-X.
3. Willett P. Similarity-based virtual screening using 2D fingerprints. *Drug Discovery Today*. 2006;11(23):1046-1053. Doi: 10.1016/j.drudis.2006.10.005.
4. Whittle M, Gillet VJ, Willett P, Loesel J. Analysis of data fusion methods in virtual screening: similarity and group fusion. *J Chem Inf Model*. 2006;46(6):2206-2219. Doi: 10.1021/ci0496144.
5. Wolpert DH. Stacked generalization. *Neural Networks*. 1992;5(2):241-259. Doi: 10.1016/S0893-6080(05)80023-1.
6. Melgar-Lesmes P, Perramon M, Jiménez W, Melgar-Lesmes P, Perramon M, Jiménez W. Roles of the hepatic endocannabinoid and apelin systems in the pathogenesis of liver fibrosis. *Cells*. 2019;8(11). Doi: 10.3390/cells8111311.
7. Yan J, Wang A, Cao J, Chen L. Apelin/APJ system: an emerging therapeutic target for respiratory diseases. *Cell Mol Life Sci*. 2020;77(15):2919-2930. Doi: 10.1007/s00018-020-03461-7.
8. Li A, Zhao Q, Chen L, Li Z. Apelin/APJ system: an emerging therapeutic target for neurological diseases. *Mol Biol Rep*. 2023;50(2):1639-1653. Doi: 10.1007/s11033-022-08075-9.
9. Yu XH, Tang ZB, Liu LJ, Qian H, Tang SL, Zhang DW, et al. Apelin and its receptor APJ in cardiovascular diseases. *Clinica Chimica Acta*. 2014;428:1-8. Doi:10.1016/j.cca.2013.09.001.
10. Hu H, He L, Li L, Chen L. Apelin/APJ system as a therapeutic target in diabetes and its complications. *Molecular Genetics and Metabolism*. 2016;119(1):20-27. Doi: 10.1016/j.ymgme.2016.07.012.
11. Luo H, Han L, Xu J. Apelin/APJ system: A novel promising target for neurodegenerative diseases. *Journal of Cellular Physiology*. 2020;235(2):638-657. Doi: 10.1002/jcp.29001.
12. Rossin D, Vanni R, Lo Iacono M, Cristallini C, Giachino C, Rastaldo R. APJ as promising therapeutic target of peptide analogues in myocardial infarction- and hypertension-induced heart failure. *Pharmaceutics*. 2023;15(5). Doi: 10.3390/pharmaceutics15051408.
13. Li C, Cheng H, Adhikari BK, Wang S, Yang N, Liu W, et al. The role of apelin-apj system in diabetes and obesity. *Front Endocrinol*. 2022;13. Doi: 10.3389/fendo.2022.820002.
14. Banerjee S, Baidya SK, Ghosh B, Adhikari N, Jha T. The first report on predictive comparative ligand-based multi-QSAR modeling analysis of 4-pyrimidinone and 2-pyridinone based APJ inhibitors. *New J Chem*. 2022;46(24):11591-11607. Doi: 10.1039/D2NJ01923J.
15. Tran XTD, Phan TL, To VT, Tran NVN, Nguyen NNS, Nguyen DNH, et al. Integration of the Butina algorithm and ensemble learning strategies for the advancement of a pharmacophore ligand-based model: an *in silico* investigation of apelin agonists. *Front Chem*. 2024;12. doi:10.3389/fchem.2024.1382319.
16. Trifonov L, Afri M, Palczewski K, Korshin EE, Gruzman A. An expedient synthesis of CMF-019: (S)-5-methyl-3-{1-(pentan-3-yl)-2-(thiophen-2-ylmethyl)-1H-benzo[d]imidazole-5-carboxamido}hexanoic acid, a potent apelin receptor (APJ) agonist. <http://www.eurekaselect.com>. Accessed December 18, 2025. <https://www.eurekaselect.com/article/89691>.
17. Ason B, Chen Y, Guo Q, Hoagland KM, Chui RW, Fielden M, et al. Cardiovascular response to small-molecule APJ activation. *JCI Insight*. 2021;5(8). doi:10.1172/jci.insight.132898.
18. Gargalovic P, Wong P, Onorato J, Finlay H, Wang T, Yan M, et al. *In vitro* and *in vivo* evaluation of a small-molecule apj (apelin receptor) agonist, BMS-986224, as a potential treatment for heart failure. *Circulation: Heart Failure*. 2021;14(3):e007351. Doi: 10.1161/circheartfailure.120.007351.
19. Truchon JF, Bayly CI. Evaluating virtual screening methods: good and bad metrics for the early recognition problem. *J Chem Inf Model*. 2007;47(2):488-508. Doi: 10.1021/ci600426e.
20. Fawcett T. An introduction to ROC analysis. *Pattern Recognition Letters*. 2006;27(8):861-874. Doi: 10.1016/j.patrec.2005.10.010.

21. Japkowicz N, He H, Ma Y. Imbalanced learning: Foundations, Algorithms, and Applications. Wiley; 2013.
22. Seidel T, Ibis G, Bendix F, Wolber G. Strategies for 3D pharmacophore-based virtual screening. Drug Discovery Today: Technologies. 2010;7(4):e221-e228. Doi: 10.1016/j.ddtec.2010.11.00.