



Ôn tập

Kiến trúc máy biến áp và cơ chế chú ý trong

# Phân tích dữ liệu bộ gen: Đánh giá toàn diện

Sanghyuk Roy Choi và Minhyeok Lee



Trường Kỹ thuật Điện và Điện tử, Đại học Chung-Ang, Seoul 06974, Hàn Quốc; choiroy@cau.ac.kr \* Thư tử: mlee@cau.ac.kr

Tóm tắt đơn giản: Lĩnh vực học sâu đang phát triển nhanh chóng, đặc biệt là các kiến trúc dựa trên máy biến áp và cơ chế chú ý, đã có khả năng ứng dụng đáng kể trong tin sinh học và phân tích dữ liệu bộ gen. Do tính chất tương tự của trình tự bộ gen với văn bản ngôn ngữ, những kỹ thuật này bước đầu thành công trong xử lý ngôn ngữ tự nhiên đã được áp dụng cho dữ liệu bộ gen.

Đánh giá này cung cấp phân tích chuyên sâu về những tiến bộ và ứng dụng gần đây nhất của các kỹ thuật này vào dữ liệu bộ gen, đánh giá nghiêm túc những ưu điểm và hạn chế của chúng. Bằng cách điều tra các nghiên cứu từ năm 2019 đến năm 2023, đánh giá này xác định các lĩnh vực nghiên cứu tiềm năng trong tương lai, từ đó khuyến khích những tiến bộ hơn nữa trong lĩnh vực này.

Tóm tắt: Sự xuất hiện và phát triển nhanh chóng của học sâu, đặc biệt là các kiến trúc dựa trên máy biến áp và cơ chế chú ý, đã có tác động mang tính biến đổi trên một số lĩnh vực, bao gồm tin sinh học và phân tích dữ liệu bộ gen. Bản chất tương tự của trình tự bộ gen với văn bản ngôn ngữ đã cho phép ứng dụng các kỹ thuật đã đạt được thành công trong các lĩnh vực xử lý ngôn ngữ tự nhiên đến dữ liệu bộ gen. Đánh giá này cung cấp một phân tích toàn diện về những tiến bộ gần đây nhất trong việc ứng dụng kiến trúc máy biến áp và cơ chế chú ý vào dữ liệu bộ gen và bản phiên mã. Trọng tâm của đánh giá này là đánh giá quan trọng các kỹ thuật này, thảo luận về những ưu điểm và hạn chế của chúng trong bối cảnh phân tích dữ liệu bộ gen. Với tốc độ phát triển nhanh chóng của các phương pháp học sâu, việc liên tục đánh giá và phản ánh về tình hình hiện tại cũng như định hướng tương lai của nghiên cứu trở nên quan trọng. Do đó, đánh giá này nhằm mục đích phục vụ như một nguồn tài nguyên kịp thời cho cả những nhà nghiên cứu dày dạn kinh nghiệm và những người mới tham gia, cung cấp cái nhìn toàn cảnh về những tiến bộ gần đây và làm sáng tỏ các ứng dụng tiềm năng trong lĩnh vực này. Hơn nữa, bài đánh giá này nhằm làm nổi bật các lĩnh vực nghiên cứu tiềm năng trong tương lai bằng cách đánh giá nghiêm túc các nghiên cứu từ năm 2019 đến năm 2023, từ đó đóng vai trò là bước đệm cho những nỗ lực nghiên cứu tiếp theo.

Từ khóa: học sâu; mô hình máy biến áp; cơ chế chú ý; dữ liệu bộ gen; dữ liệu bảng điểm; bộ gen; tin sinh học; phân tích trình tự; xử lý ngôn ngữ tự nhiên



Trích dẫn: Choi, SR; Lý, M.

Kiến trúc máy biến áp và

Cơ chế chú ý trong bộ gen

Phân tích dữ liệu: Đánh giá toàn

diện. Sinh học 2023, 12, 1033.

[https://doi.org/10.3390/](https://doi.org/10.3390/bio12071033)

bio12071033

Biên tập viên học thuật: Chi Wei Ong  
và Milan Toma

Đã nhận: ngày 20 tháng 6 năm 2023

Sửa đổi: ngày 18 tháng 7 năm 2023

Được chấp nhận: ngày 21 tháng 7 năm 2023

Đã xuất bản: ngày 22 tháng 7 năm 2023



Bản quyền: © 2023 của các tác giả.

Được cấp phép MDPI, Basel, Thụy Sĩ.

Bài viết này là một bài viết truy cập mở  
được phân phối theo các điều khoản và  
điều kiện của Creative Commons

Giấy phép ghi công (CC BY) (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Giới thiệu

Cuộc cách mạng của các phương pháp học sâu đã tiếp thêm sinh lực cho lĩnh vực tin sinh học và phân tích dữ liệu bộ gen, thiết lập nền tảng cho những tiến bộ đột phá và những hiểu biết mới lạ [1–6]. Gần đây, sự phát triển và ứng dụng các kiến trúc và cơ chế chú ý dựa trên máy biến áp đã chứng tỏ hiệu suất và khả năng vượt trội trong việc xử lý sự phức tạp vốn có của dữ liệu bộ gen. Các kỹ thuật học sâu, đặc biệt là những kỹ thuật sử dụng kiến trúc biến áp và cơ chế chú ý, đã cho thấy thành công đáng kể trong nhiều lĩnh vực khác nhau như xử lý ngôn ngữ tự nhiên (NLP) [7] và thị giác máy tính [8–10]. Những thành tựu này đã thúc đẩy việc áp dụng nhanh chóng chúng vào tin sinh học, do tính chất tương tự của trình tự bộ gen với các văn bản ngôn ngữ. Trình tự bộ gen có thể được hiểu là ngôn ngữ của sinh học và do đó, các công cụ

thành thạo trong việc xử lý dữ liệu ngôn ngữ có thể giải mã được các mẫu ẩn trong các chuỗi này.

Cơ chế chú ý, lần đầu tiên được giới thiệu trong các mô hình tuần tự [11], đã cách mạng hóa cách các mô hình học sâu xử lý và diễn giải dữ liệu [12-18]. Kỹ thuật này được thiết kế để khắc phục những hạn chế của các mô hình hồi quy truyền thống bằng cách cung cấp một cơ chế để xử lý các phần khác nhau của chuỗi đầu vào khi tạo đầu ra.

Trong bối cảnh dữ liệu bộ gen, điều này ngụ ý khả năng xem xét các vùng gen khác nhau và mối quan hệ của chúng một cách linh hoạt trong quá trình giải thích. Cơ chế chú ý tính toán tổng có trọng số của các tính năng đầu vào, trong đó trọng số, còn được gọi là điểm chú ý, được xác định động dựa trên dữ liệu đầu vào. Cơ chế này cho phép mô hình tập trung nhiều hơn vào các tính năng thiết yếu hoặc phù hợp và ít hơn vào các tính năng không liên quan hoặc ít hơn.

những cái quan trọng.

Lấy cảm hứng từ sự thành công của cơ chế chú ý, mô hình máy biến áp được đề xuất như một sự thay đổi hoàn toàn từ bản chất xử lý tuần tự của mạng thần kinh tái phát (RNN) và các biến thể của chúng [19-22]. Mô hình máy biến áp tận dụng các cơ chế chú ý để xử lý song song dữ liệu đầu vào, cho phép tính toán nhanh hơn và hiệu quả hơn. Kiến trúc của mô hình máy biến áp bao gồm một chồng các mô-đun máy biến áp giống hệt nhau, mỗi mô-đun có hai lớp con: cơ chế tự chú ý nhiều đầu và mạng cấp dữ liệu chuyển tiếp được kết nối đầy đủ theo vị trí. Sử dụng kiến trúc này, các mô hình máy biến áp có thể nắm bắt được sự phụ thuộc giữa đầu vào và đầu ra mà không cần quan tâm đến khoảng cách của chúng trong chuỗi.

Tiềm năng của kiến trúc dựa trên máy biến áp và cơ chế chú ý trong phân tích dữ liệu bộ gen là rất lớn và phần lớn chưa được khám phá. Họ trình bày một giải pháp đầy hứa hẹn để giải quyết quy mô lớn và tính chất phức tạp của dữ liệu gen. Khả năng nắm bắt sự phụ thuộc tầm xa giữa các vị trí bộ gen, xem xét đồng thời nhiều vùng gen có liên quan và tập trung thích ứng vào các đặc điểm nổi bật làm cho các phương pháp này trở nên phù hợp duy nhất cho các ứng dụng bộ gen. Bài đánh giá này tìm cách nêu bật và điều tra các ứng dụng đổi mới của các phương pháp này trong phân tích dữ liệu bộ gen, đánh giá nghiêm túc những ưu điểm và hạn chế của chúng, đồng thời đưa ra các hướng nghiên cứu trong tương lai.

Sự gia tăng nghiên cứu trong lĩnh vực này đã dẫn đến một lượng lớn các nghiên cứu và ấn phẩm, mỗi nghiên cứu đều đóng góp những phát hiện, phương pháp và quan điểm mới. Mặc dù sự phát triển nhanh chóng của nghiên cứu này là một minh chứng cho tính năng động của lĩnh vực này, nhưng nó cũng đặt ra thách thức cho các nhà nghiên cứu trong việc theo kịp những tiến bộ. Do đó, sự cần thiết phải có các tài liệu đánh giá toàn diện để quản lý, tổng hợp và trình bày một cách mạch lạc những phát hiện này là điều tối quan trọng.

Bài viết đánh giá này nhằm mục đích cung cấp một sự tổng hợp chặt chẽ và cập nhật về các tài liệu ủng hộ sự sống trong lĩnh vực này. Với tốc độ phát triển nhanh chóng của các phương pháp học sâu, điều quan trọng là phải liên tục đánh giá và phản ánh về tình hình hiện tại và hướng nghiên cứu trong tương lai. Đánh giá này sẽ đóng vai trò là nguồn tài nguyên kịp thời cho cả các nhà nghiên cứu có kinh nghiệm và những người mới tham gia lĩnh vực này, mang đến cái nhìn toàn cảnh về những tiến bộ gần đây và làm sáng tỏ các ứng dụng tiên tiến của kiến trúc máy biến áp và cơ chế chú ý trong phân tích dữ liệu bộ gen.

Đánh giá này thực hiện đánh giá có hệ thống và quan trọng đối với các nghiên cứu gần đây nhất kéo dài từ năm 2019 đến năm 2023. Việc kiểm tra kỹ lưỡng các ấn phẩm này nhằm mục đích đưa ra những quan điểm mới, phát hiện những khoảng trống nghiên cứu hiện có và đề xuất các hướng điều tra trong tương lai.

Hơn nữa, đánh giá này nhằm mục đích làm nổi bật những tác động sâu rộng và lợi ích tiềm năng liên quan đến việc áp dụng các kỹ thuật học sâu tiên tiến trong phân tích dữ liệu bộ gen. Bằng cách điều tra những tiến bộ này, nó tìm cách truyền cảm hứng và kích thích những nỗ lực nghiên cứu sâu hơn và những đột phá về công nghệ trong lĩnh vực tin sinh học năng động.

## 2. Deep Learning với Transformers và Cơ chế chú ý 2.1. Kiến trúc

thông thường của học sâu Trong những năm gần

đây, lĩnh vực y sinh đã chứng kiến sự bùng nổ đáng kể trong ứng dụng học máy và đặc biệt hơn là các phương pháp học sâu. Những quảng cáo này-

các kỹ thuật tiên tiến là công cụ giúp tìm ra những hiểu biết sâu sắc từ các bộ dữ liệu y sinh phức tạp, tạo điều kiện cho tiến bộ trong chẩn đoán bệnh, khám phá thuốc và nghiên cứu di truyền.

Học sâu, hay mạng thần kinh sâu (DNN), sử dụng mạng thần kinh nhân tạo với nhiều lớp, một tính năng giúp nó có khả năng học các mẫu phức tạp từ các tập dữ liệu lớn [23]. Một trong những dạng đơn giản nhất của mạng nơron là perceptron đa lớp (MLP), chứa một lớp đầu vào, một hoặc nhiều lớp ẩn và một lớp đầu ra. MLP thành thạo trong việc xử lý các tập dữ liệu trong đó đầu vào và đầu ra có chung mối quan hệ tuyến tính hoặc phi tuyến tính. Tuy nhiên, chúng kém hiệu quả hơn khi xử lý dữ liệu không gian hoặc thời gian, một hạn chế được khắc phục bằng các mô hình học sâu phức tạp hơn như mạng thần kinh tích chập (CNN) [24] và RNN [25].

CNN đặc biệt hiệu quả trong việc xử lý dữ liệu không gian, chẳng hạn như hình ảnh, nhờ khả năng nắm bắt các phụ thuộc cục bộ trong dữ liệu bằng cách sử dụng các lớp chập. Trong y sinh học, CNN đã chứng tỏ được công cụ hữu ích trong các nhiệm vụ như phân tích hình ảnh y tế và xác định kiểu hình mô.

RNN, bao gồm cả biến thể nâng cao của chúng, mạng bộ nhớ ngắn hạn dài (LSTM), được thiết kế để xử lý dữ liệu tuần tự bằng cách kết hợp cơ chế giống như bộ nhớ, cho phép chúng học từ các đầu vào trước đó trong chuỗi. Đặc tính này làm cho chúng có giá trị trong việc dự đoán trình tự protein hoặc hiểu trình tự di truyền trong tin sinh học.

Mạng đối thủ sáng tạo (GAN), một công cụ thay đổi trò chơi trong lĩnh vực này, bao gồm hai mạng thần kinh, bộ tạo và bộ phân biệt đối xử, cạnh tranh [26-32]. Kiến trúc độc đáo này cho phép tạo ra các phiên bản dữ liệu tổng hợp mới giống với dữ liệu huấn luyện, một tính năng đầy hứa hẹn trong khám phá thuốc và y học cá nhân hóa.

Một số biến thể khác của kỹ thuật học sâu cũng tồn tại. Ví dụ, sự chú ý đến biểu đồ tận dụng cơ chế chú ý để cân nhắc ảnh hưởng của các nút trong biểu đồ, đóng một vai trò quan trọng trong sinh học phân tử để nhận dạng cấu trúc. Mạng dư (ResNets) sử dụng kết nối phím tắt để giải quyết vấn đề biến mất độ dốc trong mạng sâu, một tính năng có thể có giá trị trong phân tích hình ảnh y tế. AdaBoost, một thuật toán tăng cường, hoạt động bằng cách kết hợp nhiều bộ phân loại yếu để tạo ra một bộ phân loại mạnh. Seq2Vec là một cách tiếp cận để xử lý dữ liệu chuỗi trong đó chuỗi được chuyển đổi thành biểu diễn vectơ có độ dài cố định. Cuối cùng, bộ mã hóa tự động biến thiên (VAE) là các mô hình tổng quát có thể tìm hiểu cách biểu diễn tiềm ẩn của dữ liệu đầu vào, mang lại tiềm năng đáng kể trong các nhiệm vụ như phát hiện bất thường hoặc giảm kích thước trong dữ liệu y sinh phức tạp.

## 2.2. Máy biến áp và cơ chế chú ý

Mô hình máy biến áp đại diện cho một bước ngoặt trong quá trình phát triển của các mô hình học sâu [33]. Khác biệt với các mô hình truyền tải trình tự thông thường, thường bao gồm các lớp lặp lại hoặc tích chập, mô hình biến áp chỉ khai thác các cơ chế chú ý, thiết lập tiền lệ mới trong các nhiệm vụ như dịch máy và xử lý ngôn ngữ tự nhiên (NLP).

Thành phần chính của mô hình máy biến áp là cơ chế chú ý và nó có hai dạng: tự chú ý (còn được gọi là chú ý nội tâm) và chú ý nhiều đầu. Chức năng cốt lõi của cơ chế chú ý là mô hình hóa các tương tác giữa các phần tử khác nhau trong một chuỗi, từ đó nắm bắt được sự phụ thuộc giữa chúng mà không quan tâm đến vị trí của chúng trong chuỗi. Về bản chất, nó quyết định mức độ cần chú ý đến các phần khác nhau của đầu vào khi tạo ra một đầu ra cụ thể.

Cơ chế tự chú ý hoạt động bằng cách tạo ra sự thể hiện của từng phần tử theo trình tự nhằm nắm bắt tác động của tất cả các phần tử khác trong trình tự. Điều này đạt được bằng cách tính điểm cho từng cặp phần tử, áp dụng hàm softmax để thu được trọng số và sau đó sử dụng các trọng số này để tạo thành tổng có trọng số của các biểu diễn phần tử ban đầu. Do đó, nó cho phép mỗi phần tử trong chuỗi tương tác với tất cả các phần tử khác, cung cấp một bức tranh tổng thể hơn về toàn bộ chuỗi.

Mặt khác, cơ chế chú ý nhiều đầu về cơ bản là nhiều cơ chế tự chú ý, hay nhiều đầu, hoạt động song song. Mỗi đầu tính toán một cách độc lập một phép biến đổi tuyến tính đã học khác nhau của đầu vào và đầu ra của chúng được nối với nhau

và biến đổi tuyến tính để tạo ra đầu ra cuối cùng. Điều này cho phép mô hình nắm bắt được nhiều loại mối quan hệ và sự phụ thuộc khác nhau trong dữ liệu.

Ngoài cơ chế tự chú ý, một khía cạnh quan trọng khác của kiến trúc máy biến áp là sự kết hợp của mã hóa vị trí. Cho rằng bản thân mô hình là bất biến hoán vị (nghĩa là nó không có bất kỳ khái niệm vốn có nào về thứ tự của các phần tử đầu vào), cần có một số phương pháp để kết hợp thông tin về vị trí của các phần tử trong chuỗi. Mã hóa vị trí phục vụ mục đích này.

Mã hóa vị trí được thêm vào phần nhúng đầu vào ở cuối ngăn xếp bộ mã hóa và bộ giải mã. Những phần nhúng này được học hoặc cố định và mục đích của chúng là đưa thông tin về vị trí tương đối hoặc tuyệt đối của các từ trong chuỗi. Việc bổ sung mã hóa vị trí cho phép mô hình tận dụng thứ tự của trình tự, điều này rất quan trọng để hiểu dữ liệu có cấu trúc như ngôn ngữ.

Một cách tiếp cận phổ biến để mã hóa vị trí là sử dụng các hàm sin và cos có tần số khác nhau. Với cách tiếp cận này, mỗi chiều của mã hóa vị trí tương ứng với một hàm sin hoặc cosin. Các hàm này có bước sóng tạo thành cấp số nhân từ  $2\pi$  đến  $10.000 \times 2\pi$ .

Một trong những ưu điểm chính của mô hình máy biến áp là khả năng xử lý các phụ thuộc tầm xa trong dữ liệu, một khía cạnh mà RNN và CNN truyền thống có thể gặp khó khăn do tính chất tuần tự của chúng. Bằng cách cho phép tất cả các phần tử trong chuỗi tương tác đồng thời, máy biến áp giảm bớt nhu cầu nén tất cả thông tin vào trạng thái ẩn có kích thước cố định, điều này thường dẫn đến mất thông tin trong chuỗi dài.

Ngoài ra, máy biến áp cũng đưa ra khái niệm mã hóa vị trí để khắc phục tình trạng thiếu thông tin vị trí vốn có trong cơ chế chú ý. Điều này rất quan trọng, đặc biệt là trong các nhiệm vụ mà thứ tự của các phần tử mang thông tin quan trọng.

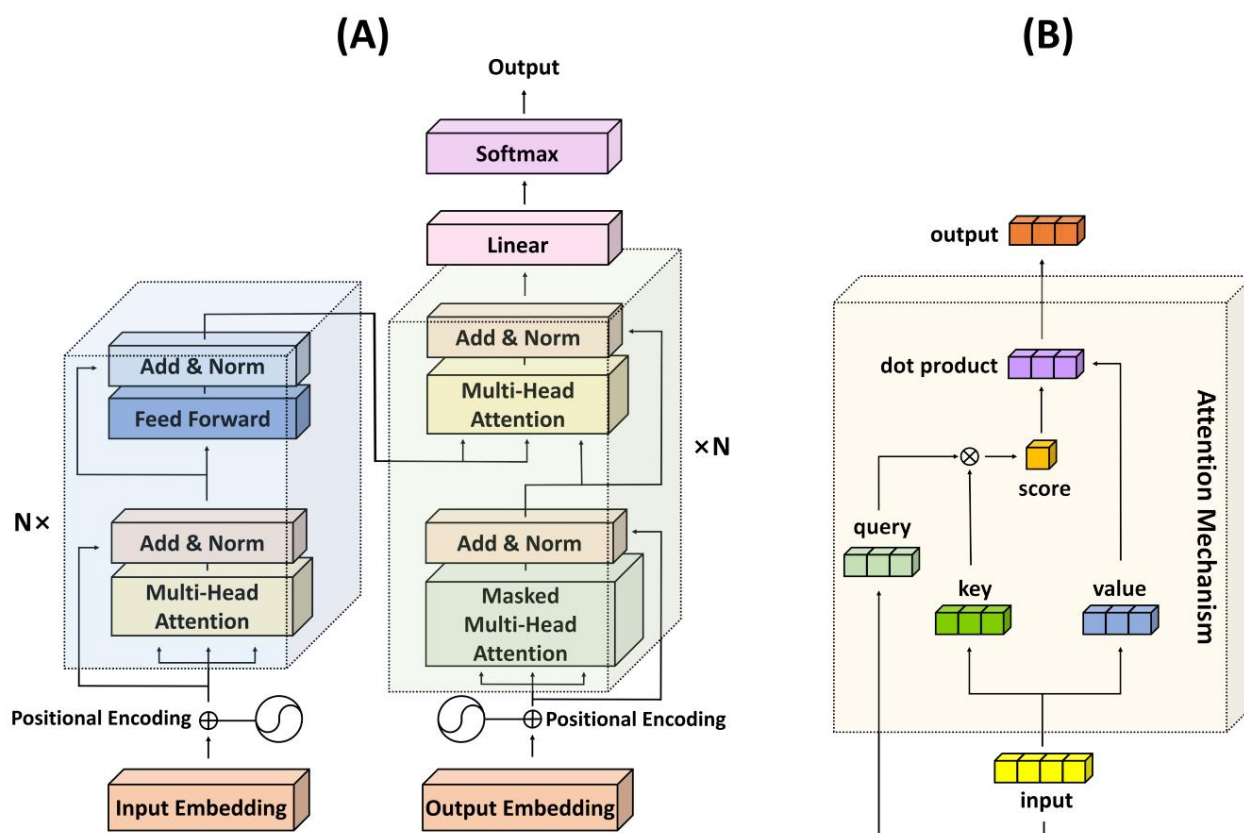
Cơ chế tự chú ý của máy biến áp bao gồm ba thành phần quan trọng: truy vấn (Q), khóa (K) và giá trị (V). Các thành phần này bắt nguồn từ các biểu diễn đầu vào và được tạo bằng cách nhân đầu vào với ma trận trọng số đã học tương ứng. Mỗi thành phần này mang một ý nghĩa riêng trong cơ chế chú ý.

Cụ thể, truy vấn tương ứng với phần tử mà chúng tôi đang cố gắng tính toán biểu diễn phụ thuộc vào ngữ cảnh. Khóa liên quan đến các phần tử mà chúng ta đang so sánh truy vấn để xác định trọng số. Cuối cùng, giá trị là phần tử được tính trọng số theo điểm chú ý (kết quả từ việc so sánh truy vấn với khóa) để tạo ra kết quả cuối cùng.

Cơ chế tự chú ý hoạt động bằng cách tính điểm chú ý cho một cặp truy vấn và khóa. Nó làm như vậy bằng cách lấy tích số chấm của họ và sau đó áp dụng hàm softmax để đảm bảo rằng các trọng số nằm trong phạm vi 0 và 1 và tổng bằng 1. Điều này cung cấp thước đo chuẩn hóa về tầm quan trọng hoặc sự chú ý mà mô hình gán cho từng phần tử khi mã hóa một phần tử cụ thể.

Sau khi tính điểm chú ý, mô hình sẽ tính tổng có trọng số của các vectơ giá trị, trong đó trọng số được tính theo điểm chú ý. Thao tác này dẫn đến mã hóa theo ngữ cảnh của từng phần tử, trong đó ngữ cảnh phụ thuộc vào tất cả các phần tử khác trong chuỗi. Những mã hóa như vậy sau đó được sử dụng làm đầu vào cho lớp tiếp theo trong mô hình máy biến áp.

Việc sử dụng ma trận Q, K và V cho phép mô hình học cách tập trung vào các khía cạnh khác nhau của dữ liệu đầu vào và cho phép mô hình phân biệt phần thông tin nào là quan trọng khi mã hóa một phần tử cụ thể. Như vậy, cơ chế chú ý của máy biến áp mang lại mức độ linh hoạt và sức mạnh đáng kể cho mô hình, cho phép nó xử lý nhiều nhiệm vụ khác nhau một cách hiệu quả và hiệu quả. Cấu trúc của kiến trúc máy biến áp và cơ chế chú ý được mô tả trong Hình 1.



Hình 1. Minh họa kiến trúc máy biến áp và cơ chế chú ý. (A) Cấu trúc máy biến áp ; (B) Cơ chế chú ý.

### 3. Phương pháp

#### 3.1. Quy trình lựa chọn xuất bản

Quá trình lựa chọn bài viết được thiết kế để đảm bảo bao gồm những đóng góp nghiên cứu chất lượng cao phù hợp với lĩnh vực trọng tâm đánh giá của chúng tôi. Để đạt được điều này, chúng tôi chủ yếu tận dụng các phương pháp tiếp cận thuật toán dựa trên công cụ tìm kiếm học thuật, Web of Science (WOS). Chúng tôi đã lựa chọn cẩn thận các từ khóa tìm kiếm của mình, tập trung vào các thuật ngữ cốt lõi như “máy biến áp học sâu”, “phương pháp chú ý”, “RNA” và “dữ liệu bộ gen”. Việc lựa chọn tỉ mỉ các từ khóa tìm kiếm này là công cụ giúp xác định các bài viết có liên quan để đưa vào đánh giá của chúng tôi.

Để trình bày tổng quan có cấu trúc về kiến trúc máy biến áp và cơ chế chú ý trong bối cảnh dữ liệu gen, chúng tôi đã phân loại các bài báo được chọn theo chủ đề nghiên cứu, tức là ứng dụng cụ thể của máy biến áp và phương pháp chú ý. Sự phân loại này nhằm mục đích góp phần mang lại sự hiểu biết toàn diện về sự giao thoa giữa học sâu với máy biến áp và cơ chế chú ý đối với dữ liệu gen. Nó nhấn mạnh sự hiểu biết về các phương pháp khác nhau được sử dụng trong lĩnh vực này. Để có bản tóm tắt ngắn gọn về các bài báo được đánh giá, hãy tham khảo Bảng 1. Chúng tôi thừa nhận rằng một số bài báo có thể phù hợp với nhiều danh mục, nhưng với mục đích đánh giá này, mỗi bài báo được phân loại theo một danh mục duy nhất nắm bắt tốt nhất chủ đề cốt lõi của bài báo.

Đánh giá của chúng tôi chỉ tập trung vào các bài báo trên tạp chí được bình duyệt, cố tình loại trừ các bản in trước và tài liệu hội nghị, mặc dù chúng có rất nhiều trong lĩnh vực này. Chúng tôi thực thi tiêu chí này để duy trì độ tin cậy và giá trị của đánh giá, từ đó đảm bảo rằng chỉ những nghiên cứu được xem xét kỹ lưỡng nghiêm ngặt mới được đưa vào. Chúng tôi mong muốn duy trì tính mới và độc đáo của bài đánh giá và do đó, cố tình loại trừ các loại bài viết cụ thể, chẳng hạn như bài viết đánh giá và quan điểm. Mục tiêu là nhấn mạnh các nghiên cứu dựa trên nghiên cứu cơ bản theo mục đích đánh giá của chúng tôi.

Bảng 1. Tổng quan về các ứng dụng của Kiến trúc máy biến áp và Cơ chế chú ý đối với dữ liệu bộ gen.

Đề tài nghiên cứu	Học
Dự đoán trình tự và địa điểm	Raad và cộng sự. [34], Shen và cộng sự. [35], Li và cộng sự. [36], Ma và cộng sự. [37], Zeng và cộng sự. [38], Mai và cộng sự. [39], Song và cộng sự. [40], Đường và cộng sự. [41], Du và cộng sự. [42], Pan và cộng sự. [43], Zhuang và cộng sự. [44], Hoàng và cộng sự. [45], Quan và cộng sự. [46], Li và cộng sự. [47], Liu và cộng sự. [48], Du và cộng sự. [49], Văn Kinh và cộng sự. [50], Cao và cộng sự. [51], Ông và cộng sự. [52], Shen và cộng sự. [53], Zhang và cộng sự. [54], Jiang và cộng sự. [55], Bhukya và cộng sự. [56], Muneer và cộng sự. [57], Wekesa và cộng sự. [58], Liang và cộng sự. [59], Zhang và cộng sự. [60], Xie và cộng sự. [61], Fan và cộng sự. [62], Tsukiyama và cộng sự. [63], Gao và cộng sự. [64], Ullah và cộng sự. [65], Quách và cộng sự. [66], Wang và cộng sự. [67], Sun và cộng sự. [68]
Biểu hiện gen và dự đoán kiểu hình	Zhang và cộng sự. [69], Lee và cộng sự. [70], Chen và cộng sự. [71], Kang và cộng sự. [72], Liao và cộng sự. [73], Angenent-Mari và cộng sự. [74], Zuo và cộng sự. [75], Karbalayghareh và cộng sự. [76], Phạm và cộng sự. [77], Dominic và cộng sự. [78], Lee và cộng sự. [79], Li và cộng sự. [80], Bu và cộng sự. [81], Schapke và cộng sự. [82]
Nghiên cứu ncRNA và CircRNA	Lưu và cộng sự. [83], Song và cộng sự. [85], Liu và cộng sự. [86], Wang và cộng sự. [87], Liu và cộng sự. [88], Gao và Shang [89], Yuan và Yang [90], Song và cộng sự. [91], Chen và cộng sự. [92], Yang và cộng sự. [93], Li và cộng sự. [94], Wang và cộng sự. [95], Li và cộng sự. [96], Fan và cộng sự. [97], Sheng và cộng sự. [98], Niu và cộng sự. [99], Zhang và cộng sự. [100], Liu và cộng sự. [101], Quách và cộng sự. [102]
Thông tin chi tiết về quá trình phiên âm	Clauwaert và cộng sự. [103], Feng và cộng sự. [104], Han và cộng sự. [105], Tao và cộng sự. [106], Asim và cộng sự. [107], Park và cộng sự. [108], Yan và cộng sự. [109], Song và cộng sự. [110]
Nhiệm vụ đa omics/phương thức	Công và cộng sự. [111], Kayikci và Khoshgoftaar [112], Ye và cộng sự. [113], Kang và cộng sự. [114], Wang và cộng sự. [115], Chan và cộng sự. [116]
Dự đoán kết quả và hiệu quả của CRISPR	Lưu và cộng sự. [117], Liu và cộng sự. [118], Wan và Jiang [119], Xiao và cộng sự. [120], Mathis và cộng sự. [121], Zhang và cộng sự. [122], Zhang và cộng sự. [123]
Suy luận mạng lưới quy định gen	Lin và Ou-Yang [124], Xu và cộng sự. [125], Feng và cộng sự. [126], Ullah và Ben-Hur [127], Xie và cộng sự. [128]
Tiên lượng bệnh	Lee [129], Choi và Lee [130], Dutta và cộng sự. [131], Xing và cộng sự. [132], Mạnh và cộng sự. [133], Feng và cộng sự. [134]
Phân loại dựa trên biểu hiện gen	Gokhale và cộng sự. [135], Beykikhoshk và cộng sự. [136], Manica và cộng sự. [137], Lee và cộng sự. [138]
proteomics	Hou và cộng sự. [139], Gong và cộng sự. [140], Armenteros và cộng sự. [141], Littmann và cộng sự. [142]
Nhận dạng loại tế bào	Bài hát và cộng sự. [143], Feng và cộng sự. [144], Buterez và cộng sự. [145], Zhang và cộng sự. [146]
Dự đoán tương tác thuốc-thuốc	Schwarz và cộng sự. [147], Kim và cộng sự. [148], Liu và Xie [149], Wang và cộng sự. [150]
Chủ đề khác	Yu và cộng sự. [151], Yamaguchi và Saito [152], Chu và cộng sự. [153], Cao và cộng sự. [154], Gupta và Shankar [155], Zhang và cộng sự. [156], Choi và Chae [157]

Chúng tôi giới hạn khoảng thời gian đánh giá của mình trong các bài viết được xuất bản từ năm 2019 đến năm 2023. Hạn chế này đảm bảo rằng đánh giá của chúng tôi vẫn diễn ra đồng thời và phù hợp, cung cấp sự hiểu biết toàn diện về những tiến bộ và xu hướng gần đây nhất trong lĩnh vực học sâu đối với dữ liệu gen. Đáng chú ý là bài đánh giá của chúng tôi chỉ tập trung vào các bài báo được bình duyệt. Quyết định này được thúc đẩy bởi hai yếu tố chính: Thứ nhất, quá trình đánh giá ngang hàng

là một cơ chế quan trọng để duy trì chất lượng và độ tin cậy của tài liệu khoa học bằng cách nghiên cứu phải được kiểm tra nghiêm ngặt bởi các chuyên gia tên miền. Thứ hai, được bình duyệt các tạp chí theo truyền thống được coi là nguồn đáng tin cậy và đáng tin cậy để xuất bản một cách khoa học nghiên cứu đúng đắn và có ảnh hưởng.

Chúng tôi đã tiến hành thu thập dữ liệu cho đến tháng 5 năm 2023, phù hợp với lịch trình hiện tại, từ đó đảm bảo rằng nội dung của đánh giá phù hợp với những phát triển mới nhất của lĩnh vực này. Trong quá trình thu thập dữ liệu, chúng tôi tổng hợp thông tin về số lượng trích dẫn và nhật ký xuất bản cho mỗi bài viết được chọn. Những dữ liệu này rất quan trọng để đánh giá phạm vi, tác động và sự chấp nhận của nghiên cứu trong cộng đồng khoa học. Chúng ta sẽ phân tích những dữ liệu này trong các phần tiếp theo của đánh giá này.

Một số nghiên cứu đã bị loại trừ dựa trên các tiêu chí cụ thể. Bài viết đánh giá không được xem xét do chúng tôi tập trung vào nghiên cứu sơ cấp. Các nghiên cứu chỉ sử dụng học máy các phương pháp không có yếu tố học sâu cũ ng bị loại trừ. Hơn nữa, giấy tờ không liên quan trực tiếp đến dữ liệu gen, chẳng hạn như dữ liệu tập trung vào phân đoạn hình ảnh, đã bị loại bỏ, mặc dù khả năng ứng dụng chung của cơ chế chú ý vào dữ liệu hình ảnh. Do đó, những nghiên cứu liên quan đến hình ảnh như vậy đã bị xóa thủ công khỏi bài đánh giá của chúng tôi.

3.2. Tạp chí của các bài báo đã xuất bản

Bảng 2 minh họa sự phân bố của các bài báo đã xuất bản tập trung vào việc áp dụng kiến trúc biến áp và cơ chế chú ý đến dữ liệu bộ gen, trên nhiều loại tạp chí khoa học.

Bảng 2. Phân bố các bài báo đã xuất bản trên các tạp chí khác nhau.

tạp chí	Đếm	Tỷ lệ phần trăm (%)
Tóm tắt về tin sinh học	20	16.1
Tin sinh học	9	7.3
Tin sinh học BMC	9	7.3
Biên giới trong di truyền học	9	7.3
Giao dịch IEEE-ACM về Sinh học tính toán và Tin sinh học	7	5,6
Sinh học tính toán PLOS	4	3.2
Truyền thông thiên nhiên	4	3.2
Khoa học liên ngành-Khoa học đời sống tính toán	4	3.2
Tạp chí Công nghệ sinh học tính toán và cấu trúc 3		2.4
Báo cáo khoa học 3		2.4
Sinh học-Basel 2		1.6
Toán sinh học và kỹ thuật 2		1.6
Biên giới trong ung thư 2		1.6
Sinh học tính toán và Hóa học 2		1.6
Kỷ yếu của Viện Hàn lâm Khoa học Quốc gia Hoa Kỳ 2		1.6
Phương pháp 2		1.6
Nghiên cứu axit nucleic 2		1.6
Tế bào 2		1.6
Biên giới trong tế bào và sinh học phát triển 2		1.6
Khác (<2 ấn phẩm) 34		27,4

Rõ ràng từ Bảng 2 là 'Tóm tắt về tin sinh học' có số lượng ấn phẩm (20), chiếm 16,1% tổng số nghiên cứu trong lĩnh vực này. 'Tin sinh học', Các tạp chí 'Tin sinh học BMC' và 'Biên giới trong di truyền' theo sát, mỗi tạp chí đều đóng góp 7,3% tổng số ấn phẩm. Các tạp chí như 'Sinh học tính toán PLOS', 'Thiên nhiên Tài khoản Truyền thông' và 'Khoa học liên ngành-Khoa học đời sống tính toán' khoảng 3,2% mỗi cái.

Hơn nữa, có một phần đáng kể các bài báo (27,4%) được phân phối ở nhiều dạng khác nhau. các tạp chí khác, mỗi tạp chí đóng góp ít hơn hai ấn phẩm. Những kết quả này thể hiện một cách rộng rãi phổ biến nghiên cứu về chủ đề này trên nhiều tạp chí khác nhau, gợi ý liên ngành

sự quan tâm và ảnh hưởng của kiến trúc máy biến áp và các ứng dụng cơ chế chú ý trong lĩnh vực phân tích dữ liệu bộ gen.

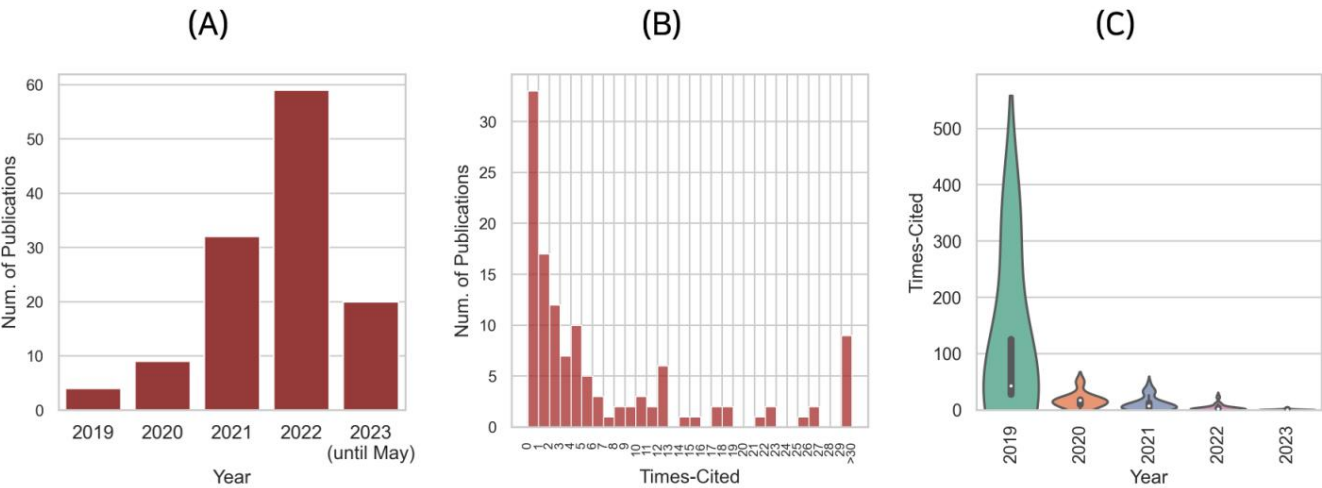
3.3. Phân tích các ấn phẩm hàng năm Như

được minh họa trong Hình 2A, xu hướng các ấn phẩm về việc sử dụng kiến trúc máy biến áp và cơ chế chú ý đến dữ liệu bộ gen cho thấy sự gia tăng đáng kể trong vài năm qua. Sự tăng trưởng và cường độ xuất bản từ năm này sang năm khác cho thấy mối quan tâm đang nổi lên nhanh chóng và hoạt động nghiên cứu chuyên sâu trong lĩnh vực này.

Năm 2019, số lượng ấn phẩm tương đối ít, chỉ có bốn nghiên cứu được ghi lại, cho thấy giai đoạn nghiên cứu mới trong lĩnh vực này. Tuy nhiên, số lượng ấn phẩm đã tăng hơn gấp đôi, lên chín ấn phẩm vào năm 2020. Điều này cho thấy sự phát triển mới nổi của lĩnh vực này khi nhiều cộng đồng nghiên cứu bắt đầu nhận ra tiềm năng biến đổi của kiến trúc máy biến áp và cơ chế chú ý để phân tích dữ liệu bộ gen.

Năm 2021 đánh dấu một bước đột phá đáng kể trong lĩnh vực này, với 32 ấn phẩm, nhiều hơn ba lần so với năm 2020. Sự gia tăng đột ngột này có thể là do sự trưởng thành của các phương pháp và sự thừa nhận ngày càng tăng về tiện ích cũng như hiệu quả của chúng trong việc giải thích dữ liệu bộ gen.

Vào năm 2022, hoạt động nghiên cứu đạt đỉnh cao với mức cao kỷ lục là 59 ấn phẩm, đánh dấu một bước ngoặt lớn và đánh dấu sự chuyển đổi của lĩnh vực này sang một giai đoạn trưởng thành hơn. Sự phổ biến của các kỹ thuật này trong phân tích dữ liệu bộ gen có thể là do khả năng sâu sắc của chúng trong việc xử lý các bộ dữ liệu bộ gen lớn và tạo ra những hiểu biết sinh học có ý nghĩa.



Hình 2. Mô hình phân phối theo năm xuất bản và tần suất trích dẫn (A) Phân bố theo năm xuất bản. (B) Phân phối tần suất trích dẫn. (C) Mối quan hệ giữa số trích dẫn và năm xuất bản.

Vào năm 2023, tính đến tháng 5, đã có 20 ấn phẩm xuất bản, cho thấy sự quan tâm mạnh mẽ đến lĩnh vực này vẫn tiếp tục. Mặc dù chỉ mới nửa năm nhưng số lượng ấn phẩm đã đạt khoảng 1/3 tổng số cho năm 2022, cho thấy động lực nghiên cứu trong lĩnh vực này dự kiến sẽ tiếp tục.

Xu hướng tăng lên về số lượng ấn phẩm trong những năm qua biểu thị sự thừa nhận và áp dụng ngày càng tăng của kiến trúc máy biến áp và các kỹ thuật cơ chế chú ý trong phân tích dữ liệu bộ gen. Nó nhấn mạnh tầm quan trọng của nghiên cứu sâu hơn để tận dụng các phương pháp học sâu đầy hứa hẹn này để giải thích sâu sắc, chính xác và nâng cao hơn về dữ liệu gen phức tạp.



### 3.4. Phân tích phân phối trích dẫn

Việc phân phối trích dẫn của các bài báo được đánh giá cung cấp dữ liệu sâu sắc về tác động học thuật và sự công nhận của chúng trong cộng đồng học thuật. Như được mô tả trong Hình 2B, minh họa biểu đồ trích dẫn và Hình 2C, thể hiện mối tương quan giữa số lượng trích dẫn và năm xuất bản của các bài báo, có một mô hình đáng chú ý trong phân bố trích dẫn.

Số lượng trích dẫn trung bình là 2 và giá trị trung bình là 9,7, cho thấy sự phân bố các trích dẫn có độ lệch dương. Độ lệch này chỉ ra rằng trong khi hầu hết các bài báo nhận được ít trích dẫn thì một số ít các bài báo được trích dẫn nhiều, điều này làm tăng đáng kể giá trị trung bình. Điều đáng chú ý là một số lượng lớn các nghiên cứu vẫn chưa được trích dẫn, chủ yếu là do chúng mới được xuất bản và chưa có đủ thời gian để xem xét. Kịch bản này nhấn mạnh tầm quan trọng của đánh giá hiện tại, nhằm mục đích cung cấp một cuộc kiểm tra kỹ lưỡng về các nghiên cứu này.

Xem xét dữ liệu trích dẫn chưa đầy đủ cho năm 2023, rõ ràng là hầu hết mọi bài báo xuất bản trong năm nay đều chưa được trích dẫn, với số lượng trích dẫn trung bình là 0. Quan sát này phù hợp với xu hướng học thuật được mong đợi trong đó các ấn phẩm mới hơn thường có ít trích dẫn hơn do độ trễ thời gian cố hữu trong quá trình trích dẫn.

Tuy nhiên, các ấn phẩm trước đó có số lượng trích dẫn cao hơn, biểu thị tác động rộng hơn và vị thế đã được thiết lập trong lĩnh vực này. Ví dụ: số lượng trích dẫn trung bình cho các bài báo được xuất bản vào năm 2019 và 2020 lần lượt là 42 và 17. Điều này cho thấy tác động học thuật đáng kể, chứng tỏ rằng chủ đề được xem xét ở đây có giá trị và sự quan tâm đáng kể đối với cộng đồng nghiên cứu.

Về vấn đề này, một số bài báo được trích dẫn nhiều đã có tác động đặc biệt đáng kể đến lĩnh vực này. Ví dụ, tác phẩm của Armenteros et al. [141], đã giới thiệu TargetP 2.0, một phương pháp tiên tiến để xác định tín hiệu phân loại đầu N trong protein bằng cách sử dụng học sâu, đã thu hút được sự chú ý đáng kể, với 333 trích dẫn cho đến nay. Lỗi chú ý của mô hình học sâu của họ nhấn mạnh rằng dư lượng thứ hai trong protein, sau methionine ban đầu, có ảnh hưởng mạnh mẽ đến việc phân loại, một đặc điểm trước đây chưa được nhấn mạnh. Điều này nhấn mạnh cách các phương pháp học sâu có thể tạo ra những hiểu biết mới về hệ thống sinh học.

Một bài báo có ảnh hưởng khác là tác phẩm của Manica et al. [137], đã đề xuất một kiến trúc mới để dự đoán có thể giải thích được độ nhạy của hợp chất chống ung thư bằng cách sử dụng bộ mã hóa tích chập dựa trên sự chú ý đa phương thức. Công trình này đã nhận được 56 trích dẫn và mô hình dự đoán của nó vượt trội đáng kể so với mô hình tiên tiến nhất trước đây về dự đoán độ nhạy của thuốc. Các tác giả cũng cung cấp một phân tích toàn diện về trọng số chú ý, chứng minh rõ hơn tính dễ hiểu của phương pháp này.

Cuối cùng, nghiên cứu của Angenent-Mari et al. [74], sử dụng học sâu để dự đoán hành vi của các phần tử RNA được thiết kế, được gọi là công tắc giữ chân, cũng nổi bật.

Với 50 trích dẫn, công trình này cho thấy DNN được đào tạo về trình tự nucleotide vượt trội hơn rất nhiều so với các mô hình trước đó dựa trên nhiệt động lực học và động học.

Những công trình được trích dẫn nhiều này nhấn mạnh tiềm năng biến đổi của các phương pháp học sâu, đặc biệt là những phương pháp tận dụng kiến trúc biến áp và cơ chế chú ý, trong việc nâng cao hiểu biết của chúng ta về hệ thống sinh học và thúc đẩy mô hình dự đoán trong y sinh. Phân phối trích dẫn phản ánh động lực tạm thời của ảnh hưởng của trường và sự công nhận ngày càng tăng của việc học sâu với kiến trúc biến áp và kỹ thuật cơ chế chú ý trong phân tích dữ liệu bộ gen. Cần phải có những đánh giá và phân tích sâu hơn về các bài báo gần đây để kích thích thảo luận và tăng khả năng hiển thị cũng như tác động của chúng trong cộng đồng học thuật.

## 4. Tổng quan về các nghiên cứu gần đây về kiến trúc máy biến áp và cơ chế chú ý đối với dữ liệu bộ gen

### 4.1. Dự đoán trình tự và địa điểm

Trong dự đoán trước miRNA, Raad et al. [34] đã giới thiệu miRe2e, một mô hình học sâu dựa trên máy biến áp. Mô hình này đã chứng minh hiệu suất được cải thiện gấp 10 lần

so với các thuật toán hiện có khi được xác nhận bằng cách sử dụng bộ gen của con người. Tương tự, Zeng et al. [38] đã giới thiệu 4mCPred-MTL, một mô hình học tập đa tác vụ kết hợp với một máy biến áp để dự đoán các vị trí 4mC trên nhiều loài. Mô hình này đã thể hiện khả năng học tính năng mạnh mẽ, nắm bắt các đặc điểm của trang web 4mC tốt hơn so với các mô tả tính năng hiện có.

Một số nghiên cứu đã tận dụng deep learning để dự đoán ưu tiên liên kết protein-RNA. Thần và cộng sự. [35] đã phát triển một mô hình dựa trên LSTM phân cấp và mạng chú ý vượt trội hơn các phương pháp khác. Du và cộng sự. [42] đã đề xuất một mạng lưới chú ý đa quy mô sâu (DeepMSA) dựa trên CNN để dự đoán các ưu tiên liên kết trình tự của các protein liên kết RNA (RBP). Pan và cộng sự. [43] đã phát triển một mô hình học sâu, CRMSNet, kết hợp CNN, ResNet và các khối tự chú ý nhiều đầu để dự đoán RBP cho các chuỗi RNA.

Công việc của Sun et al. [68] trình bày một công cụ học sâu có tên PrismNet, được thiết kế để dự đoán các tương tác RBP, không thể thiếu đối với chức năng RNA và điều hòa tế bào. Công cụ này nổi bật vì nó được xây dựng để phản ánh tính chất động và phụ thuộc vào điều kiện của các tương tác RBP-RNA, trái ngược với các công cụ hiện có chủ yếu dựa vào trình tự RNA hoặc cấu trúc RNA dự đoán. Nghiên cứu đã đề xuất PrismNet bằng cách tích hợp dữ liệu cấu trúc RNA in vivo thử nghiệm với dữ liệu liên kết RBP từ bảy loại tế bào khác nhau. Phương pháp này cho phép dự đoán chính xác liên kết RBP động trong các điều kiện tế bào khác nhau.

Một khía cạnh quan trọng giúp phân biệt PrismNet là việc áp dụng cơ chế chú ý xác định các nucleotide liên kết với RBP cụ thể về mặt tính toán. Nghiên cứu đã tìm thấy sự phong phú của các biến thể thay đổi cấu trúc (được gọi là riboSNitches) trong số các vị trí gắn kết RBP động này, có khả năng cung cấp những hiểu biết mới về các bệnh di truyền liên quan đến các liên kết RBP bị điều hòa. Do đó, PrismNet cung cấp một phương pháp để truy cập các lớp tương tác RBP-RNA dành riêng cho loại tế bào mà trước đây không thể tiếp cận được, có khả năng góp phần vào sự hiểu biết và điều trị các bệnh ở người của chúng ta. Bất chấp những ưu điểm của nó, PrismNet cũng có những hạn chế tiềm ẩn. Ví dụ, hiệu quả của PrismNet phụ thuộc rất nhiều vào chất lượng và số lượng dữ liệu cấu trúc RNA in vivo thử nghiệm và dữ liệu liên kết RBP. Sự phụ thuộc này có thể hạn chế tính hữu dụng của nó trong các tình huống mà các bộ dữ liệu mở rộng như vậy không có sẵn hoặc không đầy đủ. Hơn nữa, trong khi PrismNet sử dụng cơ chế chú ý để xác định chính xác các nucleotide liên kết với RBP, việc giải thích các điểm chú ý này trong bối cảnh sinh học có thể không đơn giản, đòi hỏi phải điều tra hoặc chuyên môn bổ sung.

Li và cộng sự. [36] đã đề xuất một mô hình học sâu tổng hợp có tên m6A-BERT-Stacking để phát hiện các vị trí m6A trong các mô khác nhau của ba loài. Các kết quả thử nghiệm đã chứng minh rằng m6A-BERT-Stacking vượt trội hơn hầu hết các phương pháp hiện có dựa trên cùng một bộ dữ liệu độc lập. Tương tự, Tang et al. [41] đã trình bày Deep6MAPred, một phương pháp học sâu dựa trên CNN và Bi-LSTM để dự đoán vị trí DNA N6-methyladenosine trên các loài thực vật.

Để công nhận người quảng bá, Ma et al. [37] đã đề xuất một thuật toán học sâu, Deep-roPre. Mô hình này cho thấy độ chính xác cao trong việc xác định vùng khởi động của sinh vật nhân chuẩn. Mai và cộng sự. [39] đã sử dụng và so sánh hiệu suất của các mô hình NLP phổ biến, bao gồm XLNET, BERT và DNABERT, để dự đoán chất xúc tiến ở vi khuẩn lam nước ngọt *Synechocystis* sp. PCC 6803 và *Synechococcus elongatus* sp. UTEX2973.

Khi dự đoán khả năng tiếp cận dung môi RNA, Huang et al. [45] đề xuất một mô hình dựa trên trình tự chỉ sử dụng dữ liệu trình tự sơ cấp. Mô hình này sử dụng các lớp chú ý đã được sửa đổi với các trường tiếp nhận khác nhau để phù hợp với cấu trúc vòng gốc của chuỗi RNA. Fan và cộng sự. [62] đã đề xuất một phương pháp tính toán mới gọi là M(2)pred để dự đoán chính xác khả năng tiếp cận dung môi của RNA. Mô hình này sử dụng mạng nơ-ron nhiều lần chụp với chiến lược trích xuất đặc điểm ngữ cảnh đa quy mô.

Để dự đoán các vị trí liên kết yếu tổ phiên mã, Bhukya et al. [56] đã đề xuất hai mô hình, PCLatt và TranAtt. Mô hình này vượt trội hơn các phương pháp tiên tiến khác như DeepSEA, DanQ, TbiNet và DeepATT trong việc dự đoán vị trí liên kết giữa các yếu tố phiên mã và trình tự DNA. Cao và cộng sự. [51] đề xuất DeepARC, một giải pháp dựa trên sự chú ý

phương pháp lai kết hợp CNN và RNN để dự đoán các vị trí liên kết yếu tố phiên mã.

Muneer và cộng sự. [57] đã đề xuất hai mô hình mạng lưới thần kinh lai sâu, đó là GCN\_GRU và GCN\_CNN, để dự đoán sự phân hủy RNA từ các chuỗi RNA. Trong dự đoán về sự thoái hóa RNA, He et al. [52] đã giới thiệu RNAdagformer, một kiến trúc mô hình để dự đoán sự thoái hóa RNA. RNAdagformer vượt trội hơn các phương pháp tốt nhất trước đây trong việc dự đoán các đặc tính thoái hóa ở độ phân giải nucleotide đối với vắc xin mRNA COVID-19.

Trong việc xác định các vị trí pseudouridine (psi), Zhuang et al. [44] đã phát triển PseUdeep, một khung học sâu để xác định các vị trí psi ở ba loài: *H. sapiens*, *S. cerevisiae* và *M. musculus*. Mô hình này sử dụng cơ chế chú ý đã được sửa đổi với các trường tiếp nhận khác nhau để phù hợp với cấu trúc vòng gốc của chuỗi RNA.

Trong dự đoán về mối liên hệ giữa miRNA và bệnh, Zhang et al. [54] đã phát triển Mạng thần kinh mã hóa-giải mã chú ý sâu (D-AEDNet) để xác định vị trí của các vị trí liên kết yếu tố phiên mã (TFBS) trong chuỗi DNA. Xie và cộng sự. [61] đã trình bày một phương pháp tính toán mới dựa trên thông tin lẫn nhau theo điểm tích cực (PPMI) và mạng chú ý để dự đoán mối liên quan giữa miRNA và bệnh (MDA), được gọi là PATMDA.

Liang và cộng sự. [59] đã phát triển một mô hình học sâu, DeepEBV, để dự đoán các trang web tích hợp virus Epstein-Barr (EBV). Mô hình này tận dụng cơ chế dựa trên sự chú ý để tự động tìm hiểu các đặc điểm bộ gen cục bộ.

Các nghiên cứu gần đây đã cho thấy mối quan tâm ngày càng tăng trong việc sử dụng các cơ chế chú ý để phân tích dữ liệu bộ gen. Các mô hình dựa trên sự chú ý đã trở nên phổ biến nhờ khả năng nắm bắt các mẫu thông tin và sự phụ thuộc lâu dài trong trình tự bộ gen. Những mô hình này đã được áp dụng cho nhiều nhiệm vụ khác nhau, bao gồm dự đoán trình tự và vị trí, dự đoán ưu tiên liên kết RNA-protein, dự đoán khả năng sống sót và xác định các yếu tố chức năng trong bộ gen. Việc sử dụng các cơ chế chú ý trong các nghiên cứu này đã chứng minh hiệu suất và độ chính xác được cải thiện, làm nổi bật tính hiệu quả của phương pháp này trong việc trích xuất thông tin có ý nghĩa từ dữ liệu bộ gen.

#### 4.2. Biểu hiện gen và dự đoán kiểu hình

Các mô hình học sâu đã được sử dụng rộng rãi để dự đoán biểu hiện gen và kiểu hình, cho thấy những cải tiến đáng kể so với các phương pháp truyền thống. Những mô hình này đặc biệt hiệu quả trong việc nắm bắt các tương tác gen-gen và gen-môi trường phức tạp và tích hợp các loại dữ liệu gen và biểu sinh đa dạng.

Một nghiên cứu đặc biệt đáng chú ý về biểu hiện gen và dự đoán kiểu hình là của Angenent-Mari et al. [74]. Công trình của họ khám phá ứng dụng của DNN để dự đoán chức năng của các công tắc giữ ngón chân, đóng vai trò là mô hình quan trọng trong sinh học tổng hợp. Những công tắc này, các phần tử RNA được thiết kế, có thể phát hiện các phân tử nhỏ, protein và axit nucleic. Tuy nhiên, việc dự đoán hành vi của chúng đã đặt ra một thách thức đáng kể—một tình huống mà Angenent-Mari và các đồng nghiệp đã tìm cách giải quyết thông qua khả năng nhận dạng mẫu nâng cao từ học sâu.

Phương pháp được các tác giả sử dụng liên quan đến việc tổng hợp và mô tả đặc điểm của bộ dữ liệu bao gồm 91.534 công tắc chân, bao gồm 23 bộ gen của virus và 906 yếu tố phiên mã ở người. Các DNN được đào tạo về các trình tự nucleotide này vượt trội đáng kể so với các mô hình động học và nhiệt động lực học tiên tiến trước đó trong việc dự đoán chức năng chuyển đổi vị trí bám chân. Hơn nữa, các tác giả đã giới thiệu hình ảnh trực quan hóa sự chú ý mà con người có thể hiểu được (VIS4Map), tạo điều kiện thuận lợi cho việc xác định các chế độ thành công và thất bại. Kiến trúc mạng bao gồm các mạng MLP, CNN và LSTM được đào tạo trên nhiều đầu vào khác nhau, bao gồm các chuỗi được mã hóa một lần và các tính năng hợp lý. Một mô hình MLP tổng hợp cũng đã được đề xuất, kết hợp cả chuỗi được mã hóa một lần và các tính năng hợp lý.

Ưu điểm của phương pháp này là rất đa dạng. Các tác giả đã tận dụng deep learning để dự đoán chức năng của các công tắc giữ chân, một nhiệm vụ trước đây đã đặt ra những thách thức đáng kể. Sự vượt trội của các mô hình tiên tiến nhất trước đây là một minh chứng cho tính hiệu quả của phương pháp được đề xuất. Hơn nữa, việc đưa vào VIS4Map sự chú ý-

trực quan hóa nâng cao khả năng diễn giải của mô hình, cung cấp những hiểu biết sâu sắc có giá trị về hoạt động của mô hình và tạo điều kiện thuận lợi cho việc xác định các lĩnh vực thành công và những lĩnh vực cần cải thiện. Mặc dù có những bước tiến đáng kể nhưng phương pháp này cũng có những hạn chế nhất định. Quá trình đào tạo đòi hỏi tính toán cao, đòi hỏi các bộ xử lý đồ họa và phần cứng có dung lượng cao mà không phải tất cả các nhà nghiên cứu đều có thể truy cập được. Hơn nữa, như với bất kỳ mô hình nào, tính tổng quát của phương pháp này đối với các loại thành phần RNA hoặc DNA khác vẫn cần được xác nhận. Cũng cần lưu ý rằng mặc dù mô hình này hoạt động tốt hơn các mô hình trước đó nhưng vẫn còn nhiều chỗ cần cải thiện vì giá trị bình phương R cao nhất đạt được là 0,70, cho thấy mô hình có thể giải thích 70% độ biến thiên của dữ liệu.

Một lĩnh vực trọng tâm chính là dự đoán biểu hiện gen dựa trên sửa đổi histone. Lee và cộng sự. [70] đã phát triển Chromoformer, một kiến trúc học sâu dựa trên máy biến áp xem xét các cửa sổ gen lớn và các tương tác nhiễm sắc thể ba chiều. Tương tự, Chen và cộng sự. [71] đã giới thiệu TransferChrome, một mô hình sử dụng mạng tích chập được kết nối dày đặc và các lớp tự chú ý để tổng hợp các tính năng toàn cầu của dữ liệu sửa đổi histone. Liao và cộng sự. [73] cũng đề xuất một mạng bộ nhớ ngắn hạn dài tích chập và hai chiều với cơ chế chú ý cho nhiệm vụ này.

Những mô hình này đã chứng minh khả năng dự đoán chính xác mức độ biểu hiện gen dựa trên tín hiệu sửa đổi histone.

Một số nghiên cứu cũng tập trung vào việc dự đoán biểu hiện gen và kiểu hình dựa trên các loại dữ liệu gen và biểu sinh khác. Ví dụ, Zhang và cộng sự. [69] đã phát triển T-GEM, một mô hình học sâu có thể giải thích được để dự đoán kiểu hình dựa trên biểu hiện gen. Kang và cộng sự. [72] đã đề xuất một mô hình học sâu dựa trên nhiều sự chú ý, tích hợp nhiều điểm đánh dấu để mô tả các cơ chế điều hòa gen phức tạp. Những mô hình này đã cho thấy khả năng tích hợp các loại dữ liệu đa dạng và nắm bắt các tương tác phức tạp, dẫn đến hiệu suất dự đoán được cải thiện.

Một số nghiên cứu cũng tập trung vào việc dự đoán các loại kiểu hình cụ thể. Ví dụ, Lee và cộng sự. [79] đã đề xuất BP-GAN, một mô hình sử dụng mạng đối thủ tổng quát (GAN) kết hợp với cơ chế chú ý để dự đoán Điểm phân nhánh RNA (BP). Những nghiên cứu này đã cho thấy tiềm năng của các mô hình học sâu trong việc dự đoán các loại kiểu hình cụ thể.

Các nghiên cứu gần đây đã tập trung vào việc sử dụng các mô hình học sâu với cơ chế chú ý để dự đoán biểu hiện gen và kiểu hình dựa trên dữ liệu gen và biểu sinh đa dạng. Những mô hình này đã cho thấy những cải tiến so với các phương pháp truyền thống bằng cách nắm bắt các tương tác gen-gen và gen-môi trường phức tạp và tích hợp các loại dữ liệu khác nhau. Cụ thể, các mô hình dựa trên sự chú ý đã được sử dụng để dự đoán mức độ biểu hiện gen bằng cách sử dụng dữ liệu sửa đổi histone, chẳng hạn như Chromoformer [70], TransferChrome [71] và mạng LSTM tích chập và hai chiều lai [73]. Ngoài ra, các nhà nghiên cứu đã khám phá khả năng dự đoán các kiểu hình cụ thể, chẳng hạn như chức năng chuyển đổi vị trí ngón chân [74] và Điểm phân nhánh RNA [79], cho thấy tính linh hoạt và tiềm năng của việc học sâu với các cơ chế chú ý trong biểu hiện gen và dự đoán kiểu hình.

#### 4.3. Nghiên cứu ncRNA và CircRNA

Việc áp dụng các mô hình học sâu, đặc biệt là các mô hình kết hợp kiến trúc máy biến áp và cơ chế chú ý, đã được khám phá rộng rãi trong nghiên cứu về RNA không mã hóa (ncRNA) và RNA tuần hoàn (circRNA). Những mô hình này đã cho thấy kết quả đầy hứa hẹn trong việc dự đoán mối liên quan giữa ncRNA với bệnh, tương tác giữa lncRNA với protein và tương tác giữa CircRNA-RBP, cùng với các nhiệm vụ khác.

Yang và cộng sự. [93] đã trình bày một phương pháp tính toán mới gọi là iCircRBP-DHN, tận dụng mạng lưới phân cấp sâu để phân biệt các vị trí gắn kết CircRNA-RBP. Cốt lõi của phương pháp này là sự kết hợp giữa mạng dư đa quy mô sâu và các đơn vị tái phát có kiểm soát hai chiều (BiGRU) được trang bị cơ chế tự chú ý. Kiến trúc này đồng thời trích xuất thông tin theo ngữ cảnh cục bộ và toàn cầu từ các chuỗi CircRNA. Nghiên cứu đã đề xuất hai sơ đồ mã hóa mới để làm phong phú thêm các biểu diễn đặc trưng.

Đầu tiên, KNFP (Mẫu tần số Nucleotide K-tuple), được thiết kế để nắm bắt các đặc điểm bối cảnh cục bộ ở nhiều quy mô khác nhau, giải quyết hiệu quả vấn đề thiếu thông tin vốn có trong cách biểu diễn một nóng thông thường. Thứ hai, CircRNA2Vec, dựa trên thuật toán Doc2Vec và nhằm mục đích nắm bắt các đặc điểm bối cảnh toàn cầu bằng cách mô hình hóa các phụ thuộc tầm xa trong các chuỗi CircRNA. Phương pháp này xử lý các chuỗi như một ngôn ngữ và ánh xạ các chuỗi con (từ) thành các vectơ phân tán, góp phần nắm bắt ngữ nghĩa và cú pháp của các chuỗi này. Hiệu quả của iCircRBP-DHN đã được xác nhận trên nhiều bộ dữ liệu CircRNA và RNA tuyến tính, đồng thời nó cho thấy hiệu suất vượt trội so với các thuật toán tiên tiến.

Mặc dù iCircRBP-DHN thể hiện một số ưu điểm nhưng nó cũng tiềm ẩn những hạn chế. Điểm mạnh của phương pháp này bao gồm khả năng mô hình hóa cả bối cảnh cục bộ và toàn cầu trong các chuỗi, khả năng chống lại sự mất ổn định về số lượng và khả năng mở rộng của nó, được thể hiện bằng hiệu suất trên các bộ dữ liệu mở rộng. Tuy nhiên, hiệu suất của phương pháp này phụ thuộc rất nhiều vào chất lượng của dữ liệu trình tự và tính hiệu quả của các sơ đồ mã hóa CircRNA2Vec và KNFP, mà có thể không nắm bắt được tất cả các sắc thái của tương tác CircRNA-RBP. Mặc dù cơ chế tự chú ý có thể cung cấp một số hiểu biết sâu sắc về những gì mô hình cho là quan trọng, nhưng nó có thể không cung cấp lời giải thích đầy đủ về lý do đằng sau những dự đoán của mô hình.

Một số nghiên cứu đã tập trung vào việc dự đoán mối liên quan giữa lncRNA và bệnh tật. Lưu và cộng sự. [83] đã phát triển một mô hình mạng chú ý kép, sử dụng hai lớp chú ý cho nhiệm vụ này, vượt trội so với một số phương pháp mới nhất. Tương tự, Gao và Shang [89] đã đề xuất một mô hình tính toán mới, DeepLDA, sử dụng DNN và cơ chế chú ý đồ thị để tìm hiểu lncRNA và những thuốc để dự đoán mối quan hệ tiềm năng giữa lncRNA và kháng thuốc. Fan và cộng sự. [97] đã đề xuất GCRFLDA, một phương pháp dự đoán mối liên hệ bệnh lncRNA-mới dựa trên việc hoàn thiện ma trận tích chập đồ thị.

Sheng và cộng sự. [98] đã phát triển VADLP, một mô hình được thiết kế để dự đoán mối liên hệ giữa lncRNA và bệnh tật bằng cách sử dụng cơ chế chú ý. Những mô hình này đã chứng tỏ khả năng dự đoán chính xác mối liên quan giữa lncRNA với bệnh, cung cấp những hiểu biết có giá trị về vai trò của lncRNA trong sự phát triển và tiến triển của bệnh.

Ngoài việc dự đoán mối liên hệ giữa lncRNA với bệnh tật, các mô hình học sâu cũng đã được sử dụng để dự đoán các tương tác giữa lncRNA với protein. Bài hát và cộng sự. [84] đã trình bày một khung học tập tổng hợp, RLF-LPI, để dự đoán các tương tác lncRNA-protein. Wekesa và cộng sự. [85] đã phát triển một phương pháp học biểu diễn đồ thị, GPLPI, để dự đoán các tương tác lncRNA-protein (LPI) thực vật từ trình tự và thông tin cấu trúc. Những mô hình này đã cho thấy khả năng nắm bắt được sự phụ thuộc giữa trình tự và cấu trúc, dẫn đến hiệu suất dự đoán được cải thiện.

Trong nhiệm vụ phân biệt RNA vòng (circRNA) với RNA không mã hóa dài khác (lncRNA), Liu et al. [101] đã đề xuất mạng học tập đa trường dựa trên sự chú ý (MIL). Mô hình này vượt trội hơn các mô hình hiện đại trong nhiệm vụ này.

Một số nghiên cứu cũng tập trung vào việc dự đoán tương tác giữa CircRNA và RBP.

Wu và cộng sự. [86] đã đề xuất một phương pháp dành riêng cho RBP, iDeepC, để dự đoán các vị trí gắn với RBP trên các CircRNA từ các chuỗi. Yuan và Yang [90] đã phát triển một phương pháp học sâu, DeCban, để xác định các tương tác giữa CircRNA và RBP. Niu và cộng sự. [99] đã đề xuất CRBPD, một mô hình tính toán sử dụng mạng phân cấp sâu tích hợp Adaboost để xác định các vị trí gắn kết của RNA-RBP vòng tròn. Những mô hình này đã chứng tỏ khả năng dự đoán chính xác các tương tác giữa CircRNA và RBP, cung cấp những hiểu biết có giá trị về vai trò của CircRNA trong quá trình điều hòa sau phiên mã. Quách và cộng sự. [102] đã đề xuất một mô hình học sâu, Circ2CBA, để dự đoán các vị trí gắn kết CircRNA-RBP. Mô hình đã đạt được giá trị AUC là 0,8987, vượt trội so với các phương pháp khác trong việc dự đoán vị trí liên kết giữa CircRNA và RBP.

Ngoài việc dự đoán các tương tác, các mô hình học sâu còn được sử dụng để dự đoán và giải thích các sửa đổi RNA sau phiên mã và các họ ncRNA.

Bài hát và cộng sự. [91] đã trình bày MultiRM, một phương pháp dự đoán và giải thích tích hợp các sửa đổi RNA sau phiên mã từ các chuỗi RNA. Chen và cộng sự. [92] đã phát triển ncDENSE, một phương pháp dựa trên mô hình học sâu để dự đoán và diễn giải

họ RNA không mã hóa từ trình tự RNA. Những mô hình này đã cho thấy khả năng dự đoán và giải thích chính xác các sửa đổi RNA và các họ ncRNA, cung cấp những hiểu biết sâu sắc có giá trị về vai trò của các sửa đổi và họ này trong việc điều hòa gen.

Một số nghiên cứu cũ ng tập trung vào việc dự đoán mối liên quan giữa CircRNA và bệnh tật. Li và cộng sự. [96] đã đề xuất một phương pháp gọi là GATGCN sử dụng mạng chú ý đồ thị và mạng đồ thị tích chập (GCN) để khám phá mối liên hệ giữa tuần hoàn RNA và bệnh tật ở người dựa trên dữ liệu đa nguồn. Vương và cộng sự. [95] đã đề xuất CDA-SKAG, một mô hình học sâu để dự đoán mối liên quan giữa tuần hoàn RNA và bệnh tật. Li và cộng sự. [94] đã giới thiệu một mô hình học sâu, GGAECDA, để dự đoán mối liên quan giữa tuần hoàn-RNA và bệnh tật. Những mô hình này đã chứng minh khả năng dự đoán chính xác mối liên quan giữa bệnh tuần hoàn với RNA, cung cấp những hiểu biết có giá trị về vai trò của CircRNA đối với sự phát triển và tiến triển của bệnh.

Các nghiên cứu gần đây đã tập trung vào việc sử dụng các mô hình học sâu với cấu trúc biến áp và cơ chế chú ý để phân tích ncRNA và CircRNA. Những mô hình này đã cho thấy nhiều hứa hẹn trong nhiều nhiệm vụ khác nhau, bao gồm dự đoán mối liên quan giữa ncRNA với bệnh , tương tác giữa lncRNA với protein, tương tác giữa CircRNA với RBP và xác định các biến đổi RNA và họ ncRNA. Việc tích hợp các cơ chế chú ý trong các mô hình này đã cải thiện độ chính xác của dự đoán và tạo điều kiện thuận lợi cho việc giải thích các tương tác và kiểu mẫu phức tạp trong dữ liệu gen.

#### 4.4. Thông tin chi tiết về quá trình phiên âm

Trong những tiến bộ gần đây, học sâu, đặc biệt là các cơ chế chú ý và mô hình chuyển đổi, đã được sử dụng đáng kể trong việc giải mã quá trình phiên mã dữ liệu bộ gen. Clauwaert và cộng sự. [103], Park và cộng sự. [108], và Han và cộng sự. [105] đã đề xuất các mô hình biến đổi tập trung vào dự đoán và mô tả đặc điểm vị trí gắn với yếu tố phiên mã (TF) .

Là một trong những ví dụ cụ thể, Yan et al. [109] đã giới thiệu một khung học sâu cải tiến để phân biệt vị trí liên kết với CircRNA-RBP, được gọi là iCircRBP-DHN, Phân biệt đối xử với các trang web liên kết với RNA-RBP dạng tuần hoàn tích hợp theo Mạng phân cấp. Họ đã giải quyết các vấn đề phổ biến với các mô hình tính toán trước đó, chẳng hạn như khả năng mở rộng kém và tính không ổn định về số, đồng thời phát triển một phương pháp biến đổi giúp hợp nhất thông tin theo ngữ cảnh cục bộ và toàn cầu thông qua mạng dư đa quy mô sâu BiGRU với cơ chế tự chú ý.

Một trong những ưu điểm chính của phương pháp này là sự kết hợp của hai sơ đồ mã hóa, CircRNA2Vec và mẫu tần số nucleotide K-tuple, cho phép biểu diễn các mức độ phụ thuộc nucleotide khác nhau, nâng cao khả năng phân biệt đối xử của các biểu diễn đặc điểm. Tính mạnh mẽ và hiệu suất vượt trội của phương pháp này đã được chứng minh thông qua thử nghiệm rộng rãi trên 37 bộ dữ liệu CircRNA và 31 bộ dữ liệu RNA tuyến tính, trong đó phương pháp này vượt trội hơn các thuật toán tiên tiến khác.

Clauwaert và cộng sự. [103] đã sử dụng khung mạng lưới thần kinh dựa trên máy biến áp để chú thích bộ gen của sinh vật nhân sơ, chủ yếu tập trung vào Escherichia coli. Nghiên cứu nhấn mạnh rằng một phần đáng kể của các tiểu đơn vị hoặc đầu chú ý của mô hình đã được điều chỉnh để xác định các yếu tố phiên mã và mô tả các vị trí gắn kết cũ ng như trình tự đồng thuận của chúng. Phương pháp này đã mở ra cơ hội tìm hiểu các yếu tố nổi tiếng và có thể mới lạ liên quan đến quá trình bắt đầu phiên mã. Đầy mạnh hơn nữa lĩnh vực dự đoán trang web gắn kết TF, Park et al. [108] đã giới thiệu TBiNet, một mô hình mạng lưới thần kinh sâu dựa trên sự chú ý, vượt trội về mặt định lượng so với các phương pháp hiện đại và chứng tỏ hiệu quả ngày càng tăng trong việc khám phá các mô típ liên kết TF đã biết. Nghiên cứu này nhằm mục đích tăng cường khả năng giải thích của các mô hình dự đoán vị trí gắn TF, một khía cạnh quan trọng để hiểu các cơ chế điều hòa gen và xác định các biến thể liên quan đến bệnh tật ở các vùng không mã hóa.

Han và cộng sự. [105] đã đề xuất MAResNet, một phương pháp học sâu kết hợp các cơ chế chú ý từ dưới lên và từ trên xuống và ResNet để dự đoán các trang web gắn kết TF. Hiệu suất mạnh mẽ của mô hình trên một tập dữ liệu thử nghiệm rộng lớn đã khẳng định lại tiềm năng của các cơ chế chú ý trong việc nắm bắt các mẫu phức tạp trong trình tự gen.

Một ứng dụng thú vị khác của deep learning được thấy trong nghiên cứu của Feng et al. [104], nơi họ đã phát triển một mô hình, PEPMAN, dự đoán các vị trí tạm dừng RNA polymerase II dựa trên dữ liệu NET-seq, là dữ liệu từ kỹ thuật thông lượng cao được sử dụng để lập bản đồ và định lượng chính xác hoạt động phiên mã mới sinh trên toàn bộ bộ gen. PEPMAN đã sử dụng các cơ chế chú ý để giải mã các đặc điểm trình tự quan trọng làm cơ sở cho việc tạm dừng Pol II. Những dự đoán trong mô hình của họ, kết hợp với các đặc điểm biểu sinh khác nhau, đã mang lại những hiểu biết sâu sắc về quá trình kéo dài phiên mã.

Về nội địa hóa RNA, Asim et al. [107] đã phát triển EL-RMLocNet, một mạng LSTM có thể giải thích được để dự đoán nội địa hóa nhiều ngăn liên quan đến RNA, sử dụng sơ đồ học tập biểu diễn thống kê GeneticSeq2Vec mới và cơ chế chú ý. Mô hình này đã vượt qua công cụ dự đoán tiên tiến hiện có để dự đoán nội địa hóa dưới lớp tế bào.

Khi dự đoán các vị trí gắn kết với RBP, Song et al. [110] đề xuất AC-Caps, một mạng lưới con nhộng dựa trên sự chú ý. Mô hình đạt hiệu suất cao, với AUC trung bình là 0,967 và độ chính xác trung bình là 92,5%, vượt qua các mô hình deep learning hiện có và tỏ ra hiệu quả trong việc xử lý dữ liệu trang web liên kết RBP quy mô lớn.

Tao và cộng sự. [106] đã trình bày một ứng dụng mới trong ung thư học; họ đã phát triển một mô hình học sâu có thể hiểu được, CITRUS, mô hình này suy ra các chương trình phiên mã được thúc đẩy bởi sự thay đổi soma trên các bệnh ung thư khác nhau. CITRUS đã sử dụng cơ chế tự chú ý để mô hình hóa tác động theo ngữ cảnh của những thay đổi soma đối với TF và các chương trình phiên mã xuôi dòng. Nó tiết lộ mối quan hệ giữa những thay đổi soma và TF, thúc đẩy các quyết định điều trị được cá nhân hóa trong ung thư chính xác.

Các mô hình học sâu với cơ chế chú ý và kiến trúc biến áp đã nổi lên như những công cụ mạnh mẽ để hiểu rõ hơn về quá trình phiên mã và giải mã dữ liệu bộ gen. Những mô hình này đã được áp dụng cho các nhiệm vụ khác nhau, chẳng hạn như dự đoán và mô tả đặc điểm của trang web gắn TF. Nhiều nghiên cứu đã đề xuất các mô hình biến đổi sử dụng các cơ chế chú ý để xác định TF, mô tả đặc điểm vị trí gắn kết của chúng và hiểu các cơ chế điều hòa gen. Ngoài ra, các mô hình học sâu đã được sử dụng để dự đoán vị trí tạm dừng RNA polymerase II, vị trí định vị RNA, vị trí gắn RBP và chương trình phiên mã do sự thay đổi soma trong bệnh ung thư. Những nghiên cứu này nêu bật tính hiệu quả của các cơ chế chú ý trong việc nắm bắt các mẫu phức tạp trong trình tự gen và cung cấp những hiểu biết có giá trị về quá trình phiên mã và điều hòa gen.

#### 4.5. Nhiệm vụ đa Omics/Phương thức

Khám phá và tích hợp dữ liệu đa omics và đa phương thức là những nhiệm vụ quan trọng trong việc tìm hiểu các hệ thống sinh học phức tạp. Các phương pháp học sâu, đặc biệt là cơ chế chú ý và mô hình biến áp, đã chứng kiến những tiến bộ và triển khai sâu sắc về mặt này. Nghiên cứu của Gong et al. [111], Kayikci và Khoshgoftaar [112], Ye và cộng sự. [113], và Wang và cộng sự. [115] đã sử dụng rộng rãi các phương pháp như vậy để phân loại dữ liệu y sinh và dự đoán bệnh.

Trong nghiên cứu của Kang et al. [114], một mô hình học sâu tổng hợp toàn diện để dự đoán tương tác giữa miRNA và lncRNA thực vật đã được đề xuất, có tên là PmliPEMG. Phương pháp này giới thiệu sự kết hợp của các tính năng phức tạp, mạng bộ nhớ ngắn hạn dài tích chập (ConvLSTM) đa quy mô và các cơ chế chú ý. Các tính năng phức tạp, được xây dựng bằng cách sử dụng các phép biến đổi phi tuyến tính của các tính năng cấu trúc và trình tự, nâng cao thông tin mẫu ở cấp độ tính năng. Bằng cách hình thành một ma trận từ vectơ đặc trưng phức tạp, các mô hình ConvLSTM được sử dụng làm mô hình cơ sở, điều này có lợi nhờ khả năng trích xuất và ghi nhớ các đặc điểm theo thời gian. Đáng chú ý, các mô hình được huấn luyện trên ba ma trận với các tỷ lệ khác nhau, nhờ đó nâng cao thông tin mẫu ở cấp độ tỷ lệ.

Một lớp cơ chế chú ý được tích hợp vào từng mô hình cơ sở, gán các trọng số khác nhau cho đầu ra của lớp LSTM. Lớp chú ý này cho phép mô hình tập trung vào thông tin quan trọng trong quá trình đào tạo. Cuối cùng, một phương pháp tập hợp dựa trên chiến lược quyết định mờ tham lam được triển khai để tích hợp ba mô hình cơ sở, cải thiện



hiệu quả và khả năng khái quát hóa. Cách tiếp cận này thể hiện những lợi thế đáng kể. Thứ nhất, việc sử dụng nâng cao thông tin đa cấp đảm bảo sự hiểu biết toàn diện hơn về dữ liệu cơ bản, tăng tính mạnh mẽ của phương pháp. Quyết định mở tham lam nâng cao hiệu quả của mô hình và khả năng khái quát hóa tổng thể.

Hơn nữa, việc áp dụng cơ chế chú ý cho phép mô hình tập trung vào các tính năng mang tính thông tin nhất, cải thiện độ chính xác dự đoán.

Công và cộng sự. [111] đã đề xuất MOADLN, một mạng học sâu chú ý đa omics, có khả năng khám phá các mối tương quan bên trong và giữa các bộ dữ liệu omics khác nhau để phân loại dữ liệu y sinh. Phương pháp này cho thấy tính hiệu quả của nó trong các nhiệm vụ phân loại dựa trên deep learning. Kayikci và Khoshgoftaar [112] đã đề xuất Chú ýDDI, một mô hình học sâu đa phương thức có kiểm soát để dự đoán ung thư vú bằng cách tích hợp dữ liệu lâm sàng, thay đổi số lượng bản sao và biểu hiện gen. Nó chứng minh tiềm năng cải thiện đáng kể trong việc phát hiện và chẩn đoán ung thư vú, cho thấy kết quả bệnh nhân tốt hơn. Ye và cộng sự. [113] đã triển khai phương pháp dự đoán gen mới bằng cách sử dụng mạng lưới thần kinh Siamese, một kiến trúc học sâu sử dụng hai nhánh có trọng số chung để so sánh và phân biệt sự giống hoặc khác nhau giữa các mẫu đầu vào, chứa mô-đun chú ý nhẹ để xác định gen gây ung thư buồng trứng. Cách tiếp cận này vượt trội hơn những cách khác về độ chính xác và hiệu quả. Tương tự, Wang và cộng sự. [115] đã đề xuất một mô hình mạng lưới thần kinh sâu tích hợp dữ liệu đa omics để dự đoán phản ứng của tế bào đối với các loại thuốc chống ung thư đã biết. Nó sử dụng lớp nhúng biểu đồ mới và lớp chú ý kết hợp hiệu quả các tính năng omics khác nhau, tính đến các tương tác của chúng.

Chan và cộng sự. [116] đã đề xuất một kiến trúc mạng lưới thần kinh sâu kết hợp dữ liệu kết nối cấu trúc và chức năng, trong đó đề cập đến việc lập bản đồ và phân tích toàn diện các kết nối thần kinh trong não, với dữ liệu đa omics để phân loại bệnh.

Họ đã sử dụng các lớp tích chập đồ thị để lập mô hình đồng thời dữ liệu Chụp cộng hưởng từ chức năng (fMRI) và Chụp ảnh kéo căng khuếch tán (DTI), là các kỹ thuật chụp ảnh thần kinh được sử dụng để đo lường sự thay đổi lưu lượng máu và mô hình khuếch tán trong não; và các lớp tích chập biểu đồ riêng biệt để mô hình hóa các bộ dữ liệu đa omics. Một cơ chế chú ý đã được sử dụng để hợp nhất các kết quả đầu ra này, làm nổi bật dữ liệu omics nào đóng góp nhiều nhất vào quyết định phân loại. Cách tiếp cận này chứng tỏ hiệu quả cao trong phân loại bệnh Parkinson bằng cách sử dụng nhiều kết hợp dữ liệu hình ảnh đa phương thức và dữ liệu đa omics.

Những nghiên cứu này nêu bật tiềm năng của các cơ chế chú ý và mô hình biến áp trong việc giải mã các hệ thống sinh học phức tạp và giải quyết các thách thức đa omics và đa phương thức trong nghiên cứu bộ gen.

#### 4.6. Dự đoán kết quả và hiệu quả của CRISPR

Hiệu quả và dự đoán kết quả của việc chỉnh sửa gen CRISPR-Cas9 đã được cải thiện đáng kể nhờ sự phát triển của các mô hình học sâu phức tạp. Một số nghiên cứu, bao gồm Liu et al. [118], Wan và Jiang [119], Xiao và cộng sự. [120], Mathis và cộng sự. [121], Zhang và cộng sự. [122], và Zhang và cộng sự. [123], đã sử dụng rộng rãi các mô hình như vậy để dự đoán kết quả chỉnh sửa CRISPR-Cas9, hiệu quả loại bỏ RNA hướng dẫn đơn (sgRNA) và các hoạt động ngoài mục tiêu, nâng cao độ chính xác của công nghệ chỉnh sửa gen.

Nghiên cứu của Zhang et al. [123] đã giới thiệu một phương pháp mới để dự đoán các hoạt động đúng mục tiêu và ngoài mục tiêu của CRISPR/Cas9 sgRNA. Họ đã đề xuất hai mô hình học sâu, CRISPR-ONT và CRISPR-OFFT, kết hợp CNN dựa trên sự chú ý để tập trung vào các yếu tố trình tự có tính quyết định nhất đối với hiệu quả của sgRNA. Những mô hình này cung cấp một số lợi thế chính. Đầu tiên, họ sử dụng lớp nhúng áp dụng mã hóa k-mer để chuyển đổi chuỗi sgRNA thành các giá trị số, cho phép CNN trích xuất các bản đồ đặc trưng. Kỹ thuật này đã được chứng minh là vượt trội hơn các phương pháp khác trong phân tích tuần tự. Thứ hai, các mô hình này sử dụng cơ chế chú ý để cải thiện cả khả năng dự đoán và khả năng diễn giải, tập trung vào các phần tử của chuỗi đầu vào là



phù hợp nhất với đầu ra. Điều này phản ánh cách các hạt nhân Cas9 hướng dẫn RNA quét bộ gen, nâng cao tính chân thực của mô hình.

Lưu và cộng sự. [118] đã trình bày Apindel, một mô hình học sâu sử dụng mô hình GloVe, một thuật toán học không giám sát được sử dụng rộng rãi để nắm bắt mối quan hệ ngữ nghĩa giữa các từ bằng cách phân tích các mẫu từ xuất hiện thống kê toàn cầu trong một kho văn bản lớn. Bằng cách tích hợp GloVe, mã hóa vị trí và mô hình học sâu nhúng BiLSTM và cơ chế chú ý, mô hình được đề xuất dự đoán kết quả chỉnh sửa CRISPR-Cas9 bằng cách nắm bắt các mối quan hệ ngữ nghĩa. Nó vượt trội hơn hầu hết các mô hình tiên tiến nhất trong dự đoán đột biến DNA và cung cấp các danh mục dự đoán chi tiết hơn. Cùng quan điểm, Wan và Jiang [119] đã giới thiệu TransCrispr, một mô hình kết hợp kiến trúc máy biến áp và CNN để dự đoán hiệu quả loại bỏ sgRNA trong hệ thống CRISPR-Cas9.

Mô hình thể hiện độ chính xác dự đoán và khả năng khái quát hóa vượt trội khi được thử nghiệm trên bảy bộ dữ liệu công khai.

Hơn nữa, Xiao và cộng sự. [120] đã đề xuất AttCRISPR, một mô hình không thời gian có thể giải thích được để dự đoán hoạt động đúng mục tiêu của sgRNA trong hệ thống CRISPR-Cas. Mô hình này kết hợp các phương pháp dựa trên mã hóa và nhúng bằng cách sử dụng chiến lược học tập tổng hợp và đạt được hiệu suất vượt trội so với các phương pháp hiện đại. Đáng chú ý, mô hình này kết hợp hai mô-đun chú ý, một không gian và một thời gian, để nâng cao khả năng diễn giải. Tương tự, Liu và cộng sự. [117] đã phát triển một mô hình học máy có thể giải thích được để dự đoán hiệu quả và tính đặc hiệu của hệ thống CRISPR-Cas.

Mathis và cộng sự. [121] đã sử dụng RNN hai chiều dựa trên sự chú ý để phát triển PRIDICT, một mô hình hiệu quả để dự đoán kết quả chỉnh sửa chính xác. Mô hình này đã chứng minh những dự đoán đáng tin cậy về những thay đổi di truyền quy mô nhỏ và nêu bật tính mạnh mẽ của PRIDICT trong việc cải thiện hiệu quả chỉnh sửa gen trên các loại tế bào khác nhau.

Phù hợp với dự đoán các hoạt động ngoài mục tiêu, Zhang et al. [122] đã trình bày một mô hình mới, CRISPR-IP, để khai thác hiệu quả thông tin cấp trình tự nhằm dự đoán các hoạt động ngoài mục tiêu trong hệ thống chỉnh sửa gen CRISPR-Cas9. Phương pháp của họ tích hợp CNN, BiLSTM và lớp chú ý, thể hiện hiệu suất vượt trội so với các mô hình hiện có.

Các nghiên cứu gần đây đã đạt được những tiến bộ đáng kể trong việc dự đoán hiệu quả và kết quả của việc chỉnh sửa gen CRISPR-Cas9 bằng mô hình học sâu. Những mô hình này đã chứng minh độ chính xác và hiệu suất vượt trội trong việc dự đoán kết quả chỉnh sửa CRISPR-Cas9, hiệu quả loại bỏ sgRNA và các hoạt động ngoài mục tiêu. Việc tích hợp các cơ chế chú ý trong các mô hình này đã cải thiện khả năng diễn giải và cung cấp những hiểu biết sâu sắc có giá trị về cơ chế chỉnh sửa gen CRISPR-Cas9.

#### 4.7. Suy luận mạng lưới quy định gen

Sự xuất hiện của học sâu đã cách mạng hóa việc suy luận các mạng điều hòa gen (GRN) từ dữ liệu giải trình tự RNA đơn bào (scRNA-seq), nhấn mạnh tiện ích của các kiến trúc máy học biến đổi như cơ chế chú ý và máy biến áp. Các nghiên cứu nổi bật, bao gồm Lin và Ou-Yang [124], Xu et al. [125], Feng và cộng sự. [126], Ullah và Ben-Hur [127], và Xie et al. [128], đã sử dụng các kiến trúc này để đưa ra các mô hình suy luận GRN, làm nổi bật hiệu suất vượt trội của chúng so với các phương pháp thông thường.

Nghiên cứu của Ullah và Ben-Hur [127] đã trình bày một mô hình mới, SATORI, để suy luận GRN. SATORI là một mô hình dựa trên Tự Attention được thiết kế để phát hiện các tương tác của phần tử điều tiết. SATORI tận dụng sức mạnh của deep learning thông qua việc kết hợp các lớp tích chập và cơ chế tự chú ý. Các lớp tích chập, được hỗ trợ bằng cách kích hoạt và gộp tối đa, xử lý các chuỗi gen đầu vào được biểu diễn thông qua mã hóa một lần. Mô hình này còn kết hợp thêm lớp RNN tùy chọn với các đơn vị bộ nhớ ngắn hạn dài để thu thập thông tin tạm thời trên toàn chuỗi.

Lớp tự chú ý nhiều đầu trong SATORI là thành phần quan trọng nhất của nó, được thiết kế để mô hình hóa các phản phụ thuộc trong chuỗi đầu vào bất kể khoảng cách tương đối của chúng. Tính năng này cho phép mô hình nắm bắt hiệu quả sự hợp tác của yếu tố phiên mã. Các

mô hình được đào tạo và đánh giá thông qua thuật toán tìm kiếm ngẫu nhiên để điều chỉnh siêu tham số và vùng dưới đường cong ROC để đo hiệu suất. Một trong những tính năng đặc biệt nhất của SATORI là khả năng xác định sự tương tác giữa các mô típ trình tự, góp phần vào khả năng diễn giải của nó. Nó sử dụng các gradient tích hợp để tính điểm phân bổ cho các họa tiết theo trình tự. Những thay đổi về những điểm số này sau khi đột biến mô típ có thể gợi ý những tương tác tiềm năng. Trong các thử nghiệm đo điểm chuẩn, SATORI đã chứng minh tỷ lệ phát hiện vượt trội của các tương tác yếu tố phiên mã đã được xác thực bằng thực nghiệm so với các phương pháp hiện có mà không cần phải xử lý hậu kỳ tốn kém về mặt tính toán.

Lin và Ou-Yang [124] đã đề xuất DeepMCL, một mô hình tận dụng phương pháp học tương phản đa góc nhìn để suy ra GRN từ nhiều nguồn dữ liệu hoặc nhiều điểm thời gian. DeepMCL đại diện cho mỗi cặp gen dưới dạng một tập hợp các hình ảnh biểu đồ và giới thiệu một mạng lưới thần kinh tích chập Xiêm sâu với sự mất tương phản, một hàm mất tương phản thường được sử dụng trong các nhiệm vụ học tập không giám sát hoặc tự giám sát nhằm khuyến khích các mẫu tương tự đến gần hơn trong không gian nhúng trong khi đẩy các mẫu khác nhau các mẫu cách xa nhau hơn; điều này cho phép thu được khả năng nhúng chiều thấp cho từng cặp gen. Hơn nữa, một cơ chế chú ý đã được sử dụng để tích hợp các phần nhúng được trích xuất từ các nguồn dữ liệu khác nhau và các cặp gen lân cận.

Tương tự, Xu và cộng sự. [125] đã trình bày STGRNS, một phương pháp dựa trên biến áp có thể hiểu được để suy ra GRN từ dữ liệu scRNA-seq. Phương pháp này tận dụng kỹ thuật mô-đun biểu hiện gen để chuyển đổi các cặp gen thành các vectơ phụ liên kết, sau đó đóng vai trò là đầu vào cho bộ mã hóa máy biến áp. Hơn nữa, Feng và cộng sự. [126] đã giới thiệu scGAEGAT, một mô hình đa phương thức tích hợp bộ mã hóa đồ thị tự động và mạng chú ý đồ thị để phân tích RNA-seq đơn bào, cho thấy hiệu suất đầy hứa hẹn trong việc cắt bỏ gen và dự đoán phân cụm tế bào.

Xie và cộng sự. [128] đề xuất MVIFMDA, một phương pháp tổng hợp thông tin đa góc nhìn để dự đoán mối liên hệ giữa miRNA và bệnh. Mô hình này sử dụng các mạng được xây dựng từ các mối liên hệ giữa miRNA và bệnh đã biết cũ ng như các điểm tương đồng giữa miRNA và bệnh, được xử lý bằng mạng tích chập đồ thị, theo sau là chiến lược chú ý để hợp nhất cách biểu diễn cấu trúc liên kết và biểu diễn thuộc tính.

Ứng dụng thành công của deep learning—đặc biệt là các cơ chế chú ý và mô hình biến đổi—trong suy luận GRN nêu bật tiềm năng của nó trong việc nâng cao độ chính xác của các dự đoán mạng lưới điều hòa gen và các phân tích di truyền khác. Những mô hình này đã chứng tỏ hiệu suất vượt trội và khả năng giải thích, vượt trội so với các phương pháp thông thường và cung cấp những hiểu biết có giá trị về cơ chế điều hòa gen và bệnh tật.

#### 4.8. Tiên lượng bệnh

Các mô hình học sâu với kiến trúc máy biến áp và cơ chế chú ý đã được sử dụng đáng kể trong việc ước tính tiên lượng bệnh, chứng minh tính hiệu quả của chúng trong việc trích xuất các mẫu có ý nghĩa từ dữ liệu gen phức tạp. Trong số các nghiên cứu tiên phong trong lĩnh vực này có những nghiên cứu được thực hiện bởi Lee [129], Choi và Lee [130], Dutta et al. [131], Xing và cộng sự. [132], và Mạnh và cộng sự. [133].

Lee [129] đã giới thiệu Mạng tập hợp chú ý gen (GAENET), một mô hình được thiết kế để ước tính tiên lượng bệnh u thần kinh đệm cấp độ thấp (LGG). GAENET đã kết hợp một cơ chế chú ý gen được thiết kế riêng cho dữ liệu biểu hiện gen, vượt trội hơn các phương pháp truyền thống và xác định HILS1 là gen tiên lượng quan trọng nhất đối với LGG. Tương tự, Choi và Lee [130] đã đề xuất Multi-PEN, một mô hình học sâu sử dụng các sơ đồ đa omics và đa phương thức để tiên lượng LGG. Mô hình này kết hợp các lớp chú ý gen cho từng loại dữ liệu, chẳng hạn như mRNA và miRNA, để xác định các gen tiên lượng, cho thấy hiệu suất mạnh mẽ so với các mô hình hiện có.

Sức mạnh của sự chú ý đến bản thân đã được Dutta et al. [131] thông qua mô hình đa phương thức sâu của họ, DeePROG, được thiết kế để dự báo tiên lượng của các gen bị ảnh hưởng bởi bệnh tật từ dữ liệu omics không đồng nhất. DeePROG vượt trội hơn các mô hình cơ sở trong việc trích xuất các tính năng có giá trị từ từng phương thức và tận dụng khả năng tiên lượng của dữ liệu y sinh. Mặt khác, Xing và cộng sự. [132] đã phát triển MLA-GNN, một biểu đồ chú ý đa cấp

mạng lưới thần kinh để chẩn đoán và tiên lượng bệnh. Mô hình của họ đã định dạng dữ liệu omics thành đồ thị đồng biểu thức và các tính năng đồ thị đa cấp được xây dựng, đạt được đặc biệt hiệu suất trên dữ liệu phiên mã từ bộ dữ liệu The Cancer Genome Atlas (TCGA- LGG/TCGA-GBM) và dữ liệu proteomic từ huyết thanh bệnh nhân COVID-19/không phải COVID-19.

Trong một bối cảnh khác biệt nhưng có liên quan, Meng et al. [133] đã giới thiệu một khuôn khổ mới gọi là SAVAE-Cox để phân tích tỷ lệ sống sót của dữ liệu bảng điểm chiều cao. Người mẫu kết hợp một cơ chế chú ý mới và tận dụng tối đa sự chuyển giao đối nghịch chiến lược học tập, vượt trội so với các mô hình phân tích sinh tồn tiên tiến nhất về chỉ số phù hợp. Phong và cộng sự. [134] đã áp dụng mô hình học sâu với cơ chế chú ý. Bộ phân loại có thể dự đoán chính xác tỷ lệ sống sót, với khu vực bên dưới máy thu hoạt động đường cong đặc tính (ROC) và ROC phụ thuộc thời gian đạt 0,968 và 0,974 trong tập huấn luyện tương ứng.

Kết hợp lại với nhau, những nghiên cứu này nêu bật tiềm năng của các cơ chế chú ý trong việc cải thiện ước tính tiên lượng bệnh, báo trước một mô hình mới trong phân tích dữ liệu gen cho mục đích tiên lượng. Hiệu quả của chúng đối với nhiều loại bệnh và phương thức dữ liệu biểu thị một con đường đầy hứa hẹn cho nghiên cứu trong tương lai về y học chính xác.

#### 4.9. Phân loại dựa trên biểu hiện gen

Việc triển khai các mô hình học sâu với kiến trúc máy biến áp và cơ chế chú ý đã cải thiện đáng kể độ chính xác phân loại dựa trên biểu hiện gen, như đã trình bày trong nhiều nghiên cứu của Gokhale et al. [135], Beykikhoshk và cộng sự. [136], Manica và cộng sự. [137], và Lee và cộng sự. [138].

Gokhale và cộng sự. [135] đưa ra GeneViT, một phương pháp biến đổi tầm nhìn, một phương pháp sâu kiến trúc học tập áp dụng các nguyên tắc tự chú ý và mô hình biến đổi để dữ liệu trực quan để phân loại biểu hiện gen ung thư. Cách tiếp cận sáng tạo này bắt đầu với bước giảm kích thước bằng cách sử dụng bộ mã hóa tự động xếp chồng, sau đó là bước cải tiến Thuật toán DeepInsight, là một phương pháp chuyển đổi dữ liệu không phải hình ảnh để sử dụng cho kiến trúc mạng nơ-ron tích chập, đạt được lợi thế hiệu suất vượt trội so với các phương pháp hiện có, như được quan sát từ các đánh giá trên mười bộ dữ liệu chuẩn.

Tương tự, trong nỗ lực cải thiện việc phân loại phân nhóm ung thư vú, Beykikhoshk et al. [136] đã giới thiệu DeepTRIAGE. Kiến trúc deep learning này áp dụng cơ chế chú ý để lấy được điểm số dấu ấn sinh học được cá nhân hóa, từ đó phân bổ cho mỗi bệnh nhân những thông tin có thể hiểu được và điểm số dấu ấn sinh học cá nhân. Đáng chú ý, DeepTRIAGE đã phát hiện ra một lỗ hổng đáng kể mối liên quan giữa tính không đồng nhất trong điểm số dấu ấn sinh học ở vùng thất lưng và giai đoạn khối u.

Trong một ứng dụng khác, Manica et al. [137] đã tạo ra một kiến trúc mới lạ cho dự đoán có thể giải thích được về độ nhạy của hợp chất chống ung thư. Mô hình này sử dụng bộ mã hóa tích chập dựa trên sự chú ý đa phương thức và quản lý để vượt xa cả đường cơ sở mô hình được đào tạo về dấu vân tay Morgan, một loại kỹ thuật dấu vân tay phân tử được sử dụng trong tin hóa học để mã hóa thông tin cấu trúc của các phân tử và lựa chọn bộ mã hóa dựa trên Hệ thống nhập dòng đầu vào phân tử đơn giản hóa (SMILES), cùng với trước đây đã báo cáo các phương pháp tiên tiến nhất để dự đoán độ nhạy của thuốc đa phương thức.

Lee và cộng sự. [138] đã phát triển một mô hình học sâu dựa trên lộ trình cải tiến với một cơ chế chú ý và truyền bá mạng lưới để phân loại phân nhóm ung thư. Các mô hình kết hợp các mạng tích chập đồ thị để biểu diễn từng con đường và một mô hình tập hợp dựa trên nhiều sự chú ý đã được sử dụng để hợp nhất hàng trăm con đường. Các mô hình đã chứng minh độ chính xác phân loại cao trong các thí nghiệm với năm bộ dữ liệu ung thư TCGA và tiết lộ các con đường và chức năng sinh học cụ thể của từng loại phụ, cung cấp những hiểu biết sâu sắc về các cơ chế sinh học làm nền tảng cho các phân nhóm ung thư khác nhau.

Những nghiên cứu này nêu bật tính hiệu quả và ứng dụng sáng tạo của cơ chế chú ý trong phân tích dữ liệu gen, mang lại những hiểu biết mới về y học chính xác và ung thư.

#### 4.10. proteomics

Việc sử dụng deep learning, đặc biệt là việc kết hợp các kiến trúc biến áp và cơ chế chú ý trong nghiên cứu hệ protein, đã dẫn đến những phát triển mang tính đột phá.

trong dự đoán chức năng của protein, như được mô tả trong nghiên cứu của Hou et al. [139], Gong và cộng sự. [140], Armenteros và cộng sự. [141], và Littmann và cộng sự. [142].

Hou và cộng sự. [139] đã xây dựng iDeepSubMito, một mô hình mạng lưới thần kinh sâu được thiết kế để dự đoán quá trình định vị protein trong màng sụn. Mô hình này sử dụng một lớp nhúng biểu đồ sáng tạo để đồng hóa dữ liệu tương tác làm thông tin trước đó để dự đoán. Ngoài ra, một lớp chú ý đã được kết hợp để tích hợp các tính năng omics khác nhau trong khi xem xét sự tương tác của chúng. Hiệu quả của mô hình này được xác nhận nhờ hiệu suất vượt trội so với các phương pháp tính toán khác trong quá trình xác thực chéo trên hai bộ dữ liệu chứa protein từ bốn ngân ty thể.

Trong khi đó, Gong và cộng sự. [140] đã đề xuất một thuật toán, iDRO, nhằm mục đích tối ưu hóa trình tự mRNA dựa trên trình tự axit amin nhất định của protein mục tiêu. Thuật toán của họ bao gồm một quy trình gồm hai bước bao gồm tối ưu hóa khung đọc mở (ORF) và tạo vùng chứa được dịch (UTR). Bước trước sử dụng BiLSTM-CRF để xác định codon cho từng axit amin, trong khi bước sau liên quan đến RNA-Bart để tạo ra UTR tương ứng. Các trình tự gen ngoại sinh được tối ưu hóa bằng áp dụng mô hình trình tự gen nội sinh của con người và các trình tự mRNA được tối ưu hóa bằng phương pháp của chúng cho thấy biểu hiện protein cao hơn so với các phương pháp truyền thống.

Armenteros và cộng sự. [141] đã giới thiệu TargetP 2.0, một mô hình học máy tiên tiến giúp xác định các tín hiệu sắp xếp đầu N trong peptide bằng cách sử dụng học sâu. Mô hình của họ nhấn mạnh vai trò quan trọng của dư lượng thứ hai trong việc phân loại protein, tiết lộ các kiểu phân bố độc đáo giữa các nhóm protein khác nhau và các peptide mục tiêu.

Littmann và cộng sự. [142] đã giới thiệu bindEmbed21, một phương pháp dự đoán dư lượng protein liên kết với các ion kim loại, axit nucleic hoặc các phân tử nhỏ. Mô hình này tận dụng các phần nhúng từ Mô hình ngôn ngữ protein dựa trên biến áp ProtT5, vượt trội so với các dự đoán dựa trên MSA khi sử dụng các chuỗi đơn. Suy luận dựa trên tính tương đồng đã cải thiện hơn nữa hiệu suất và phương pháp này đã tìm thấy các dư lượng liên kết trong hơn 42% tổng số protein của con người trước đây không được ngụ ý là có liên kết. Những nghiên cứu này chứng minh tiềm năng đáng kể của kiến trúc chuyển đổi và cơ chế chú ý trong các mô hình học sâu để dự đoán chức năng protein chính xác.

#### 4.11. Nhận dạng loại tế bào

Trong các nghiên cứu gần đây, việc áp dụng kiến trúc biến áp và cơ chế chú ý trong học sâu đã mang lại tiến bộ đáng kể trong việc xác định loại tế bào, chứng tỏ hiệu suất vượt trội trên nhiều loại tế bào, loài và độ sâu trình tự khác nhau.

Việc áp dụng kiến trúc máy biến áp và cơ chế chú ý trong học sâu để nhận dạng loại tế bào đã có những tiến bộ đáng kể, như đã được chứng minh trong các nghiên cứu của Song và cộng sự. [143], Feng và cộng sự. [144], Buterez và cộng sự. [145], và Zhang và cộng sự. [146].

Bài hát và cộng sự. [143] đã phát triển TransCluster, một cấu trúc mạng lai tận dụng phân tích phân biệt tuyến tính và một máy biến áp được sửa đổi để tăng cường khả năng học hỏi tính năng trong các bản đồ phiên mã đơn ô. Phương pháp này vượt trội hơn các kỹ thuật đã biết trên các bộ dữ liệu tế bào khác nhau từ các mô khác nhau của con người, cho thấy độ chính xác và độ tin cậy cao.

Phong và cộng sự. [144] đã đề xuất một mô hình mạng lưới thần kinh đồ thị có hướng có tên là scDGAE để phân tích dữ liệu RNA-seq đơn bào. Bằng cách sử dụng bộ mã hóa tự động đồ thị và mạng chú ý đồ thị, scDGAE đã giữ lại các thuộc tính kết nối của đồ thị có hướng và mở rộng trường tiếp nhận của hoạt động tích chập. Mô hình này đã xuất sắc trong việc cắt bỏ gen và dự đoán phân cụm tế bào trên bốn bộ dữ liệu scRNA-seq với nhãn tế bào tiêu chuẩn vàng.

Hơn nữa, Buterez và cộng sự. [145] đã giới thiệu CellVGAE, một quy trình làm việc để phân tích scRNA-seq không giám sát bằng cách sử dụng mạng chú ý đồ thị. Kiến trúc bộ mã hóa tự động đồ thị đa dạng này hoạt động trực tiếp trên kết nối di động để giảm kích thước và phân cụm. Vượt trội hơn cả kỹ thuật thần kinh và phi thần kinh, CellVGAE cung cấp khả năng diễn giải bằng cách phân tích các hệ số chú ý của biểu đồ, nắm bắt thời gian giả và động lực kích hoạt NF-kappa B.

Zhang và cộng sự. [146] đã giới thiệu RefHiC, một khung học sâu dựa trên sự chú ý để chú thích các cấu trúc tô pô từ Hi-C, đây là một kỹ thuật gen để đo lường

tổ chức không gian ba chiều của chất nhiễm sắc trong nhân. Bằng cách sử dụng bảng tham chiếu gồm các bộ dữ liệu Hi-C, RefHiC đã thể hiện hiệu suất vượt trội trên các loại tế bào, loài và độ sâu trình tự khác nhau.

#### 4.12. Dự đoán tương tác thuốc-thuốc

Các nghiên cứu gần đây đã cho thấy sự tiến bộ đáng chú ý trong việc dự đoán tương tác thuốc-thuốc (DDI) thông qua việc sử dụng các mô hình học sâu kết hợp cơ chế chú ý và kiến trúc máy biến áp, vượt qua các phương pháp học sâu cổ điển và khác trong khi nêu bật các cấu trúc thuốc quan trọng. Học sâu với kiến trúc máy biến áp và cơ chế chú ý đã nâng cao đáng kể khả năng dự đoán DDI.

Schwarz và cộng sự. [147] đã giới thiệu Chú ý DDI, một mạng lưới thần kinh đa phương thức tự chú ý của Xiêm tích hợp các biện pháp tương tự khác nhau của thuốc bắt nguồn từ đặc điểm của thuốc. Nó thể hiện hiệu suất cạnh tranh so với các mô hình DDI hiện đại trên nhiều bộ dữ liệu chuẩn. Tương tự, Kim và cộng sự. [148] đã phát triển DeSIDE-DDI, một khung kết hợp các dấu hiệu biểu hiện gen do thuốc gây ra để dự đoán DDI. Mô hình này xuất sắc với AUC là 0,889 và Khu vực thu hồi chính xác (AUPR) là 0,915, vượt qua các phương pháp hàng đầu khác trong dự đoán tương tác không nhìn thấy.

Hơn nữa, Liu và Xie [149] đã đề xuất TranSynergy, một mô hình tăng cường biến áp dựa trên kiến thức và tự chú ý để dự đoán sự kết hợp thuốc hiệp đồng.

TranSynergy vượt trội hơn các phương pháp hiện có và tiết lộ những con đường mới liên quan đến những sự kết hợp này, cung cấp những hiểu biết mới về y học chính xác và các liệu pháp chống ung thư. Vương và cộng sự. [150] cũng đã phát triển một mô hình học sâu, DeepDDS, để xác định sự kết hợp thuốc hiệu quả cho các tế bào ung thư cụ thể. Nó vượt qua các phương pháp học máy cổ điển và các phương pháp dựa trên học sâu khác, làm nổi bật các cấu trúc hóa học quan trọng của thuốc. Cùng với nhau, những nghiên cứu này nêu bật tính hữu ích của kiến trúc máy biến áp và cơ chế chú ý trong việc dự đoán tương tác thuốc-thuốc, mở đường cho những tiến bộ hơn nữa trong lĩnh vực này.

#### 4.13. Các chủ đề

Khác Kiến trúc máy biến áp và cơ chế chú ý đã tìm thấy ứng dụng trong các chủ đề nghiên cứu bộ gen khác nhau, nêu bật tính linh hoạt của kiến trúc máy biến áp và cơ chế chú ý trong nghiên cứu bộ gen. Ví dụ, Yu và cộng sự. [151] đã phát triển IDMIL-III, một phương pháp học tập đa trường sâu không cân bằng, dự đoán xuất sắc các tương tác isoform-isoform trên toàn bộ gen, đồng thời Yamaguchi và Saito [152] nâng cao khả năng dự đoán hiệu ứng biến thể dựa trên biến áp bằng cách đề xuất kiến trúc miền (DA)-nhận biết các giao thức tinh chỉnh tiến hóa, là các phương pháp tính toán tận dụng các thuật toán tiến hóa và xem xét các đặc điểm cấu trúc của các miền protein để tối ưu hóa và tinh chỉnh việc sắp xếp chuỗi protein.

Mặt khác, Chu và cộng sự. [153] kết hợp mạng lưới thần kinh tích chập với các máy biến thể trong mô hình học sâu, INTERACT, để dự đoán tác động của các biến thể di truyền lên mức độ methyl hóa DNA. Cao và cộng sự. [154] đã trình bày DeepASmRNA, một mô hình mạng lưới thần kinh tích chập dựa trên sự chú ý, cho thấy kết quả đầy hứa hẹn để dự đoán các sự kiện ghép nối thay thế.

Gupta và Shankar [155] đã đề xuất một cách sáng tạo miWords, một hệ thống xử lý bộ gen như các câu gồm các từ, để xác định các vùng tiền miRNA trên bộ gen thực vật, đạt được độ chính xác ấn tượng là 98%. Đồng thời, Zhang và cộng sự. [156] đã phát triển iLoc-miRNA, một mô hình học sâu sử dụng BiLSTM với khả năng tự chú ý nhiều đầu để dự đoán vị trí của miRNA trong tế bào, cho thấy tính chọn lọc cao đối với miRNA ngoại bào.

Choi và Chae [157] đã giới thiệu moBRCA-net, một khung phân loại ung thư vú, giúp cải thiện đáng kể hiệu suất bằng cách tích hợp nhiều bộ dữ liệu omics. Những nghiên cứu này cho thấy tính linh hoạt và tiềm năng của kiến trúc máy biến áp và cơ chế chú ý trong bối cảnh nghiên cứu bộ gen đa dạng.

## 5. Thảo luận

Khi xem xét các tài liệu hiện có, rõ ràng là các mô hình học sâu sử dụng kiến trúc biến áp và cơ chế chú ý đã cho thấy kết quả đầy hứa hẹn trong việc phân tích dữ liệu bộ gen. Tuy nhiên, thách thức vẫn tồn tại và cơ hội cho công việc trong tương lai rất đa dạng.

### 5.1. Thách thức

Một trong những thách thức chính vốn có trong việc áp dụng các mô hình học sâu vào dữ liệu gen liên quan đến cấu trúc phức tạp của những dữ liệu này. Cụ thể, dữ liệu biểu hiện gen thường được biểu diễn dưới dạng vectơ chiều cao do số lượng gen thu được trong mỗi mẫu trong quá trình giải trình tự thông lượng cao. Cách trình bày này đặt ra thách thức đối với các phương pháp giải thích và phân tích dữ liệu thông thường. Mặc dù một số nghiên cứu, chẳng hạn như nghiên cứu của Lee et al. [70] và Chen và cộng sự. [71], đã có những bước tiến trong khía cạnh này bằng cách đề xuất các kiến trúc mô hình mới hoặc các kỹ thuật tiền xử lý, bản chất chiều cao của dữ liệu gen vẫn còn là một thách thức.

Một thách thức đáng kể khác là tính sẵn có hạn chế của dữ liệu được dán nhãn. Trong nhiều nhiệm vụ như dự đoán mối liên hệ giữa lncRNA và bệnh tật hoặc tương tác tuần hoàn giữa RNA và RBP, số lượng mối liên hệ tích cực và tiêu cực được xác nhận bằng thực nghiệm thường không đủ để đào tạo các mô hình học sâu [83,86]. Điều này có thể dẫn đến các mô hình thiên về tầng lớp đa số và do đó mang lại hiệu quả kém cho tầng lớp thiểu số.

Sự phức tạp vốn có của các hệ thống sinh học cũng đặt ra những thách thức đáng kể. Ví dụ, các tương tác gen-gen và gen-môi trường rất phức tạp và thường phi tuyến tính, khiến chúng khó nắm bắt được bằng các mô hình học sâu tiêu chuẩn [72,74]. Hơn nữa, dữ liệu về gen và biểu sinh thường không đồng nhất, bao gồm các loại dữ liệu đa dạng như dữ liệu trình tự, dữ liệu biểu hiện gen và dữ liệu sửa đổi histone.

Việc tích hợp các loại dữ liệu đa dạng này trong một mô hình thống nhất có thể là một thách thức.

### 5.2. Công việc tương lai

Một hướng đầy hứa hẹn cho công việc trong tương lai là phát triển các kiến trúc mô hình mới có thể xử lý hiệu quả tính chất chiều cao của dữ liệu gen. Điều này có thể liên quan đến việc thiết kế các mô hình có thể tự động trích xuất các tính năng liên quan từ dữ liệu hoặc tận dụng các kỹ thuật như giảm kích thước hoặc lựa chọn tính năng. Hơn nữa, việc kết hợp kiến thức sinh học có sẵn vào thiết kế của các mô hình này có thể giúp hướng dẫn quá trình trích xuất đặc điểm và dẫn đến các mô hình dễ hiểu hơn.

Cũng cần có những phương pháp có thể giải quyết hiệu quả tình trạng sẵn có hạn chế của dữ liệu được dán nhãn trong bộ gen. Một cách tiếp cận đầy hứa hẹn là tận dụng các kỹ thuật học không giám sát hoặc bán giám sát, có thể sử dụng dữ liệu chưa được gắn nhãn để cải thiện hiệu suất mô hình [158-160]. Học chuyển, trong đó một mô hình được đào tạo trên một tập dữ liệu lớn được tinh chỉnh trên một tập dữ liệu nhỏ hơn, dành riêng cho nhiệm vụ, cũng có thể là một cách tiếp cận đầy hứa hẹn để giải quyết tình trạng khan hiếm dữ liệu được gắn nhãn [161-163].

Giải quyết sự phức tạp của các hệ thống sinh học có thể liên quan đến việc phát triển các mô hình có thể nắm bắt được các tương tác phức tạp và các mối quan hệ phi tuyến tính điển hình trong các hệ thống sinh học. Những mô hình này cần có khả năng đáp ứng tính không đồng nhất của dữ liệu gen và biểu sinh. Công việc gần đây của Kang et al. [72] và Liao và cộng sự. [73] chỉ ra tiềm năng của các mô hình học sâu đa phương thức trong vấn đề này. Cần nghiên cứu sâu hơn để phát triển và cải tiến các mô hình như vậy cho các nhiệm vụ khác nhau trong bộ gen.

Ngoài ra, việc kết hợp kiến thức lĩnh vực vào các mô hình có thể là một cách tiếp cận đầy hứa hẹn khác. Bằng cách kết hợp các cơ chế hoặc mối quan hệ sinh học đã biết vào các mô hình, chúng tôi có thể hướng dẫn quá trình học tập và làm cho các biểu diễn đã học trở nên dễ hiểu hơn.

Cuối cùng, sự xuất hiện của các mô hình dựa trên máy biến áp, chẳng hạn như dòng GPT, mang đến cơ hội thú vị cho công việc trong tương lai. Những mô hình này đã cho thấy nhiều hứa hẹn trong xử lý ngôn ngữ tự nhiên và khả năng mô hình hóa các mối phụ thuộc tầm xa, trong đó các yếu tố gen ở xa thường tương tác với nhau, có thể mang lại lợi ích lớn cho gen.

Sinh học 2023, 12, 1033

23 trên 29

Do đó, việc điều chỉnh và áp dụng các mô hình dựa trên máy biến áp này vào dữ liệu bộ gen là một hướng đi đầy hứa hẹn cho công việc trong tương lai.

6. Kết luận

Trong bối cảnh phát triển nhanh chóng của tin sinh học, nhu cầu tổng hợp toàn diện những phát triển và phương pháp luận gần đây nhất là rất cần thiết. Đánh giá này nhằm mục đích cung cấp một cuộc kiểm tra sâu rộng về cách sử dụng mạng tính biến đổi của học sâu, đặc biệt là các kiến trúc biến đổi và cơ chế chú ý, trong phân tích các tương tác protein-protein. Sự phát triển nhanh chóng của các chiến lược tính toán này đã nâng cao đáng kể khả năng xử lý và giải mã dữ liệu gen phức tạp của chúng tôi, đánh dấu một kỷ nguyên mới trong lĩnh vực này.

Phân tích được trình bày trong tài liệu này, được rút ra từ các nghiên cứu gần đây nhất từ năm 2019 đến năm 2023, nhấn mạnh tính linh hoạt đáng kinh ngạc và hiệu suất vượt trội của các kỹ thuật học sâu này trong vô số ứng dụng. Từ dự đoán trình tự và vị trí, biểu hiện gen và dự đoán kiểu hình, đến các nhiệm vụ đa omics phức tạp hơn và ước tính tiên lượng bệnh, các kỹ thuật học sâu đã chứng minh tiềm năng của chúng trong việc làm sáng tỏ các mô hình và mối quan hệ ẩn trong trình tự gen. Hơn nữa, việc áp dụng kiến trúc máy biến áp và cơ chế chú ý không chỉ đầy nhanh quá trình tính toán mà còn cải thiện độ chính xác và khả năng diễn giải, cuối cùng thúc đẩy lĩnh vực này phát triển.

Bất chấp những tiến bộ và thành công đáng chú ý đã được ghi nhận, điều quan trọng cần lưu ý là việc tích hợp deep learning vào phân tích dữ liệu bộ gen vẫn còn ở giai đoạn sơ khai. Vẫn còn một số thách thức và hạn chế cần giải quyết, đặc biệt là trong việc cải thiện khả năng diễn giải của các mô hình này và điều chỉnh chúng để sử dụng với các bộ dữ liệu nhỏ hơn, thường gặp trong lĩnh vực gen. Hơn nữa, với độ phức tạp và quy mô ngày càng tăng của dữ liệu gen, nhu cầu về các công cụ tính toán hiệu quả và tiên tiến hơn ngày càng tăng.

Thông qua đánh giá này, chúng tôi hy vọng sẽ cung cấp một nền tảng cho các nhà nghiên cứu tham gia vào những tiến bộ mới nhất, làm quen với các ứng dụng tiên tiến nhất và xác định những khoảng trống tiềm năng cũ ng như cơ hội để khám phá trong tương lai. Sự tổng hợp này, bao gồm một loạt các chủ đề nghiên cứu và ứng dụng, cho thấy tiềm năng to lớn và khả năng ứng dụng rộng rãi của các kỹ thuật học sâu trong tin sinh học.

Việc tích hợp các phương pháp học sâu, đặc biệt là kiến trúc biến áp và cơ chế chú ý, vào bộ công cụ tin sinh học đã hỗ trợ rất nhiều cho sự hiểu biết của chúng ta về 'ngôn ngữ sinh học'. Những kỹ thuật tính toán mạnh mẽ này đã được chứng minh là tài sản vô giá trong việc làm sáng tỏ những bí ẩn được mã hóa trong trình tự gen. Khi biên giới nghiên cứu này tiếp tục mở rộng và phát triển, chúng tôi dự đoán rằng những hiểu biết sâu sắc được cung cấp bởi tổng quan này sẽ thúc đẩy sự đổi mới và khám phá liên tục, thúc đẩy chúng tôi hướng tới những khám phá mới trong thế giới năng động của phân tích dữ liệu bộ gen.

Đóng góp của tác giả: Khái niệm hóa, SRC và ML; phân tích chính thức, SRC; điều tra, SRC; viết-chuẩn bị bản thảo gốc, SRC và ML; viết-đánh giá và chỉnh sửa, SRC và ML; trực quan hóa, SRC và ML; giám sát, ML Tất cả các tác giả đã đọc và đồng ý với phiên bản đã xuất bản của bản thảo.

Tài trợ: Công trình này được hỗ trợ bởi Quỹ nghiên cứu quốc gia Hàn Quốc (NRF) do chính phủ Hàn Quốc (MSIT) tài trợ (số RS-2023-00251528).

Tuyên bố của Ban Đánh giá Thẻ chế: Không áp dụng.

Tuyên bố đồng ý sau khi được thông báo: Không áp dụng.

Tuyên bố về tính sẵn có của dữ liệu: Không có dữ liệu mới nào được tạo hoặc phân tích trong nghiên cứu này.

Xung đột lợi ích: Các tác giả khẳng định không có xung đột lợi ích.



2. Lee, M. Kỹ thuật học sâu với dữ liệu gen trong tiên lượng ung thư: Đánh giá toàn diện về tài liệu giai đoạn 2021-2023. *Sinh học* 2023, 12, 893. [\[Tham khảo chéo\]](#)

3. Gomes, R.; Paul, N.; Anh, N.; Huber, AF; Jansen, RJ Ứng dụng lựa chọn tính năng và học sâu để dự đoán ung thư Sử dụng dấu hiệu methyl hóa DNA. *Gen* 2022, 13, 1557. [\[Tham khảo chéo\]](#)

4. Sadad, T.; Aurangzeb, RA; Safran, M.; Imran; Alfarhood, S.; Kim, J. Phân loại các loại virus có tính phân kỳ cao từ trình tự DNA/RNA bằng cách sử dụng mô hình dựa trên máy biến áp. *Y sinh học* 2023, 11, 1323. [\[CrossRef\]](#)

5. Lee, M. Những tiến bộ gần đây trong học sâu để phân tích tương tác protein-protein: Đánh giá toàn diện. *Phân tử* 2023, 28, 5169. [\[Tham khảo chéo\]](#)

6. Kim, Y.; Lee, M. Phương pháp tiếp cận học tập sâu cho các cơ chế qua trung gian lncRNA: Đánh giá toàn diện về sự phát triển gần đây-ment. *Int. J. Mol. Khoa học.* 2023, 24, 10299. [\[Tham khảo chéo\]](#)

7. Nâu, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, JD; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Các mô hình ngôn ngữ là những người học ít lần. *Khuyến cáo. Thần kinh Inf. Quá trình. Hệ thống.* 2020, 33, 1877-1901.

8. Khan, S.; Naseer, M.; Hayat, M.; Zamir, SW; Khan, FS; Shah, M. Transformers trong tầm nhìn: Một cuộc khảo sát. *Máy tính ACM. Sổng sót. (CSUR)* 2022, 54, 1-41. [\[Tham khảo chéo\]](#)

9. Lưu, Z.; Lin, Y.; Cao, Y.; Huh.; Ngụy, Y.; Trương, Z.; Lin, S.; Guo, B. Máy biến áp Swin: Máy biến áp tầm nhìn phân cấp sử dụng các cửa sổ được dịch chuyển. Trong *Kỷ yếu của Hội nghị Quốc tế IEEE/CVF về Thị giác Máy tính*, Montreal, BC, Canada, ngày 11-17 tháng 10 năm 2021; trang 10012-10022.

10. Hàn, K.; Vương, Y.; Chen, H.; Chen, X.; Quách, J.; Lưu, Z.; Tăng, Y.; Tiêu, A.; Xu, C.; Xu, Y.; et al. Một cuộc khảo sát về máy biến áp tầm nhìn. *IEEE Trans. Mẫu Hậu Môn. Mach. Trí tuệ.* 2022, 45, 87-110. [\[Tham khảo chéo\]](#)

11. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, AN; Kaiser, Ł.; Polosukhin, I. Sự chú ý là tất cả những gì bạn cần. *Khuyến cáo. Thần kinh Inf. Quá trình. Hệ thống.* 2017, 30, 6000-6010.

12. Ngụy, Z.; Yên, Q.; Lữ, X.; Zheng, Y.; Mặt trời, S.; Lin, J. Mạng lưới tái thiết nén với khả năng tự chú ý phối hợp và Mô-đun lọc Gaussian thích ứng. *Toán* 2023, 11, 847. [\[CrossRef\]](#)

13. Jin, A.; Zeng, X. Một phương pháp học sâu mới để nhận dạng mục tiêu dưới nước dựa trên dây thần kinh phức tạp có độ phân giải dày đặc Mạng có cơ chế chú ý. *J. Mar. Khoa học. Anh.* 2023, 11, 69. [\[Tham khảo chéo\]](#)

14. Gao, L.; Ngô, Y.; Dương, T.; Trương, X.; Tăng, Z.; Chan, CKD; Chen, W. Nghiên cứu về phân loại và truy xuất hình ảnh bằng cách sử dụng Deep Learning với cơ chế chú ý về di sản kiến trúc Trung Quốc hải ngoại ở Giang Môn, Trung Quốc. *Tòa nhà* 2023, 13, 275. [\[Tham khảo chéo\]](#)

15. Lữ, J.; Ren, H.; Thạch, M.; Thôi, C.; Trương, S.; Emam, M.; Li, L. Một phương pháp phân chia tế bào lai mới dựa trên mạng kết hợp tính năng đa quy mô và mạng lưới chú ý kép. *Điện tử* 2023, 12, 979. [\[CrossRef\]](#)

16. Cheng, S.; Liu, Y. Nghiên cứu về nhận dạng phương thức vận tải dựa trên sự kết hợp tạm thời của sự chú ý của nhiều đầu Mạng. *Cảm biến* 2023, 23, 3585. [\[CrossRef\]](#)

17. Kasgari, AB; Safavi, S.; Nouri, M.; Hâu, J.; Sarshar, NT; Ranjbarzadeh, R. Mô hình ưu tiên điểm quan tâm sử dụng sự chú ý Cơ chế trong một mạng lưới thần kinh tích chập. *Kỹ thuật sinh học* 2023, 10, 495. [\[Tham khảo chéo\]](#)

18. Raimundo, A.; Pavia, Nhật Bản; Sebastião, P.; Postolache, O. YOLOX-Ray: Trình phát hiện đối tượng một giai đoạn dựa trên sự chú ý hiệu quả Được thiết kế riêng cho việc kiểm tra công nghiệp. *Cảm biến* 2023, 23, 4681. [\[CrossRef\]](#)

19. Kim, T.; Pak, W. Phát hiện xâm nhập mạng dựa trên học tập sâu bằng cách sử dụng nhiều bộ chuyển đổi hình ảnh. ứng dụng. *Khoa học.* 2023, 13, 2754. [\[Tham khảo chéo\]](#)

20. Phùng, S.; Chu, X.; Mẹ, S.; Lan, Q. GIT: Mô hình học sâu dựa trên máy biến áp để đảo ngược địa âm. *J. Mar. Khoa học. Anh.* 2023, 11, 1108. [\[Tham khảo chéo\]](#)

21. Giang, D.; Shi, G.; Lý, N.; Mẹ, L.; Lý, W.; Shi, J. TRFM-LS: Phương pháp học sâu dựa trên máy biến áp để dự đoán quỹ đạo tàu. *J. Mar. Khoa học. Anh.* 2023, 11, 880. [\[Tham khảo chéo\]](#)

22. Cao, L.; Vương, Q.; Hồng, J.; Hàn, Y.; Trương, W.; Chung, X.; Che, Y.; Có thể.; Du, K.; Ngô, D.; et al. MVI-TR: Mô hình học sâu dựa trên máy biến áp với CT tăng cường độ tương phản để dự đoán trước phẫu thuật về sự xâm lấn vi mạch trong ung thư biểu mô tế bào gan. *Cự giải* 2023, 15, 1538. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

23. Shrestha, A.; Mahmood, A. Đánh giá các thuật toán và kiến trúc học sâu. *Truy cập IEEE* 2019, 7, 53040-53065. [\[Tham khảo chéo\]](#)

24. Bhatt, D.; Patel, C.; Talsania, H.; Patel, J.; Vaghela, R.; Pandya, S.; Modi, K.; Các biến thể Ghayvat, H. CNN cho thị giác máy tính: Lịch sử, kiến trúc, ứng dụng, thách thức và phạm vi tương lai. *Điện tử* 2021, 10, 2470. [\[CrossRef\]](#)

25. Yu, Y.; Sáu.; Hu, C.; Zhang, J. Đánh giá về mạng thần kinh tái phát: tế bào LSTM và kiến trúc mạng. *Máy tính thần kinh.* 2019, 31, 1235-1270. [\[Tham khảo chéo\]](#)

26. Lee, M.; Seok, J. Mạng đối thủ tạo ra có thể kiểm soát được. *Truy cập IEEE* 2019, 7, 28158-28169. [\[Tham khảo chéo\]](#)

27. Kim, J.; Lee, M. Tối ưu hóa danh mục đầu tư bằng cách sử dụng các mạng đối nghịch tạo ra bộ phân loại phụ trợ dự đoán. *Anh. ứng dụng. Nghệ thuật. Trí tuệ.* 2023, 125, 106739. [\[Tham khảo chéo\]](#)

28. Lee, M.; Seok, J. Mạng lưới đối thủ tổng hợp được hướng dẫn bằng điểm. *Tiên đề* 2022, 11, 701. [\[Tham khảo chéo\]](#)

29. Lee, M.; Seok, J. Ước tính với sự không chắc chắn thông qua các mạng đối nghịch tạo ra có điều kiện. *Cảm biến* 2021, 21, 6194. [\[CrossRef\]](#) [\[PubMed\]](#)

30. Yeom, T.; Lee, M. DuDGAN: Cải thiện GAN có điều kiện lớp thông qua Khuếch tán kép. *arXiv* 2023, arXiv:2305.14849.

31. Ko, K.; Lee, M. ZIGNeRF: Trình bày cảnh 3D không cảnh quay với Trường bức xạ thần kinh tạo ra có thể đảo ngược. *arXiv* 2023, arXiv:2306.02741.



32. Lee, M. Những tiến bộ gần đây trong Mạng đối thủ sáng tạo cho dữ liệu biểu hiện gen: Đánh giá toàn diện. *toán học* 2023, 11, 3055. [\[Tham khảo chéo\]](#)

33. Ngưu, Z.; Chung, G.; Yu, H. Đánh giá về cơ chế chú ý của học sâu. *Điện toán thần kinh* 2021, 452, 48–62. [\[Tham khảo chéo\]](#)

34. Raad, J.; Bugnon, LA; Milone, DH; Stegmayer, G. miRe2e: Một mô hình sâu toàn diện dựa trên máy biến áp để dự đoán của tiền miRNA. *Tin sinh học* 2022, 38, 1191–1197. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

35. Thần, Z.; Trương, Q.; Hàn, K.; Huang, DS Một mô hình học sâu để dự đoán mức độ ưu tiên liên kết với protein-RNA dựa trên LSTM phân cấp và Mạng chú ý. *Chuyển đổi IEEE-ACM. Máy tính. Biol. Thông tin sinh học.* 2022, 19, 753–762. [\[Tham khảo chéo\]](#)

36. Lý, Q.; Trình, X.; Bài hát, C.; Liu, T. M6A-BERT-Stacking: Công cụ dự đoán cụ thể cho mô để xác định RNA N6-Methyladenosine Các trang web dựa trên BERT và chiến lược xếp chồng. *Đối xứng* 2023, 15, 731. [\[Tham khảo chéo\]](#)

37. Ma, ZW; Triệu, JP; Thiên, J.; Zheng, CH DeeProPre: Công cụ dự đoán yếu tố thúc đẩy dựa trên học sâu. *Máy tính. Biol. Chem.* 2022, 101, 107770. [\[Tham khảo chéo\]](#)

38. Tăng, R.; Cheng, S.; Liao, M. 4mCPred-MTL: Xác định chính xác các vị trí DNA 4mC ở nhiều loài bằng cách sử dụng phương pháp học sâu đa tác vụ dựa trên Cơ chế chú ý nhiều đầu. *Đăng trước. Nhà phát triển di động Biol.* 2021, 9, 664669. [\[Tham khảo chéo\]](#)

39. Mai, DHA; Nguyễn, LT; Lee, EY TSSNote-CyaPromBERT: Phát triển một nền tảng tích hợp để dự đoán và trực quan hóa chất xúc tiến có độ chính xác cao về *Synechococcus* sp. và *Synechocystis* sp. thông qua mô hình xử lý ngôn ngữ tự nhiên tiên tiến BERT. *Đăng trước. Genet.* 2022, 13, 1067562. [\[Tham khảo chéo\]](#)

40. Bài hát, J.; Thiên, S.; Yu, L.; Dương, Q.; Xing, Y.; Trương, C.; Đại, Q.; Duan, X. MD-MLI: Dự đoán tương tác miRNA-lncRNA bằng cách sử dụng nhiều tính năng và học sâu phân cấp. *Chuyển đổi IEEE-ACM. Máy tính. Biol. Thông tin sinh học.* 2022, 19, 1724–1733. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

41. Tăng, X.; Zheng, P.; Lý, X.; Ngô, H.; Ngụy, ĐQ; Lưu, Y.; Huang, G. Deep6mARed: Phương pháp học sâu dựa trên CNN và Bi-LSTM để dự đoán vị trí DNA N6-methyladenosine trên các loài thực vật. *Phương pháp* 2022, 204, 142–150. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

42. Du, B.; Lưu, Z.; Luo, F. Mạng lưới chú ý đa quy mô sâu để dự đoán protein gắn với RNA. *Thông tin Khoa học.* 2022, 582, 287–301. [\[Tham khảo chéo\]](#)

43. Pan, Z.; Chu, S.; Zou, H.; Lưu, C.; Tăng, M.; Lưu, T.; Wang, Q. CRMSNet: Một mô hình học sâu sử dụng khối tích chập và khối tự chú ý nhiều đầu còn lại để dự đoán RBP cho chuỗi RNA. *Protein-Cấu trúc. Chức năng. Thông tin sinh học.* 2023, 91, 1032–1041. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

44. Trang, J.; Lưu, D.; Lin, M.; Thu, W.; Lưu, J.; Chen, S. PseUdeep: Nhận dạng địa điểm RNA Pseudouridine bằng thuật toán học sâu. *Đăng trước. Genet.* 2021, 12, 773882. [\[Tham khảo chéo\]](#)

45. Hoàng, Y.; Lạc, J.; Kinh, R.; Li, M. Phân tích dự đoán đa mô hình về khả năng tiếp cận dung môi RNA dựa trên phần dư đã sửa đổi cơ chế chú ý Ngắn gọn. *Thông tin sinh học.* 2022, 23, bbac470. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

46. Quan, X.; Vương, Y.; Thiệu, W.; Lý, Z.; Hoàng, S.; Zhang, D. S2Snet: Học sâu để nhận dạng RNA trọng lượng phân tử thấp bằng nanopore. *Ngắn gọn. Thông tin sinh học.* 2022, 23, bbac098. [\[Tham khảo chéo\]](#)

47. Lý, X.; Trương, S.; Shi, H. Một mạng dự được cải tiến bằng cách sử dụng phản ứng tổng hợp sâu để xác định các vị trí RNA 5-methylcytosine. *Tin sinh học* 2022, 38, 4271–4277. [\[Tham khảo chéo\]](#)

48. Phi, Y.; Trương, H.; Vương, Y.; Lưu, Z.; Liu, Y. LTPConstraint: Phương pháp toàn diện dựa trên học tập chuyển giao cho RNA thứ cấp dự đoán cấu trúc *Thông tin sinh học BMC.* 2022, 23, 1–26. [\[Tham khảo chéo\]](#)

49. Du, Z.; Tiểu, X.; Uversky, VN DeepA-RBPBS: Mạng nơ-ron tích chập và hồi quy kết hợp với sự chú ý cơ chế dự đoán vị trí gắn kết RBP. *J. Sinh khối. Cấu trúc. Dyn.* 2022, 40, 4250–4258. [\[Tham khảo chéo\]](#)

50. Văn Kinh, Y.; Bảo Ngọc, Z.; Tối thiểu, Z.; Thanh Xuyên, Z.; Hồng, W.; Da, M. Chú ýSplice: Sự tự chú ý của nhiều đầu có thể giải thích được Mô hình học sâu kết hợp dựa trên dự đoán trang Splice. *Cải cảm. J. Điện tử.* 2022, 31, 870–887.

51. Cao, L.; Lưu, P.; Chen, J.; Đặng, L. Dự đoán các trang web liên kết yếu tố phiên mã bằng cách sử dụng phương pháp học sâu kết hợp. *Đăng trước. Oncol.* 2022, 12, 893520. [\[Tham khảo chéo\]](#)

52. Anh ấy, S.; Cao, B.; Sabnis, R.; Sun, Q. RNAdegformer: Dự đoán chính xác sự suy giảm mRNA ở độ phân giải nucleotide bằng phương pháp học sâu. *Ngắn gọn. Thông tin sinh học.* 2023, 24, bbac581. [\[Tham khảo chéo\]](#)

53. Thần, LC; Lưu, Y.; Bài hát, J.; Yu, DJ SResNet: Mạng dự lượng tự chú ý để dự đoán liên kết DNA-protein. *Ngắn gọn. Thông tin sinh học.* 2021, 22, bbab101. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

54. Trương, Y.; Vương, Z.; Tăng, Y.; Chu, J.; Zou, Q. Dự đoán các trang web liên kết yếu tố phiên mã có độ phân giải cao đã cải thiện hiệu suất và khả năng diễn giải bằng phương pháp học sâu. *Ngắn gọn. Thông tin sinh học.* 2021, 22, bbab273. [\[Tham khảo chéo\]](#)

55. Giang, JY; Ju, CJT; Hào, J.; Chen, M.; Wang, W. JEDI: Dự đoán RNA tròn dựa trên bộ mã hóa mối nối và tương tác sâu giữa các vị trí mối nối. *Tin sinh học* 2021, 37, I289–I298. [\[Tham khảo chéo\]](#)

56. Bhukya, R.; Kumari, A.; Dasari, CM; Amilpur, S. Mạng lưới thần kinh sâu lai dựa trên sự chú ý để xác định chính xác vị trí liên kết yếu tố phiên mã. *Máy tính thần kinh. Ứng dụng.* 2022, 34, 19051–19060. [\[Tham khảo chéo\]](#)

57. Muneer, A.; Fati, SM; Akbar, NA; Agustriawan, D.; Wahyudi, ST iVaccine-Deep: Dự đoán về vắc xin mRNA COVID-19 suy thoái bằng cách sử dụng deep learning. *J. King Saud-Univ.-Comput. Thông tin Khoa học.* 2022, 34, 7419–7432. [\[Tham khảo chéo\]](#)

58. Wekesa, JS; Mạnh, J.; Luan, Y. Sự kết hợp đa tính năng để học sâu nhằm dự đoán tương tác lncRNA-protein thực vật. *Bộ gen* 2020, 112, 2928–2936. [\[Tham khảo chéo\]](#)

59. Lương, J.; Cui, Z.; Ngô, C.; Yu, Y.; Thiên, R.; Xie, H.; Jin, Z.; Quạt, W.; Xie, W.; Hoàng, Z.; et al. DeepEBV: Một mô hình học sâu để dự đoán các trang web tích hợp virus Epstein-Barr (EBV). *Tin sinh học* 2021, 37, 3405–3411. [\[Tham khảo chéo\]](#)

60. Trương, H.; Fang, J.; Mặt trời, Y.; Xie, G.; Lin, Z.; Gu, G. Dự đoán mối liên hệ giữa miRNA và bệnh tật thông qua biểu đồ chú ý cấp độ nút Bộ mã hóa tự động. *Chuyển đổi IEEE-ACM. Máy tính. Biol. Thông tin sinh học.* 2023, 20, 1308–1318. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

61. Tạ, X.; Vương, Y.; Anh ấy, K.; Sheng, N. Dự đoán mối liên quan giữa miRNA và bệnh dựa trên PPMI và mạng lưới chú ý. *Thông tin sinh học BMC.* 2023, 24, 1–19. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

62. Quạt, XQ; Hu, J.; Đường, YX; Giả, NX; Yu, DJ; Zhang, GJ Dự đoán khả năng tiếp cận dung môi RNA từ tính năng bối cảnh đa quy mô thông qua mạng lưới thần kinh multi-shot. *Hậu môn. Hóa sinh.* 2022, 654, 114802. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

63. Tsukiyama, S.; Hasan, MM; Đặng, HW; Kuzata, H. BERT6mA: Dự đoán vị trí DNA N6-methyladenine bằng phương pháp sâu phương pháp tiếp cận dựa trên học tập. Ngắn gọn. *Thông tin sinh học.* 2022, 23, bbac053. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

64. Cao, Y.; Chen, Y.; Phong, H.; Trương, Y.; Yue, Z. RicENN: Dự đoán chất tăng cường giống lúa bằng mạng lưới thần kinh dựa trên DNA Trình tự. *Liên ngành-Khoa học-Máy tính. Khoa học cuộc sống.* 2022, 14, 555–565. [\[Tham khảo chéo\]](#)

65. Ullah, A.; Malik, KM; Saudagar, AKJ; Khan, MB; Abul Hasanat, MH; ALTameem, A.; AlKhatami, M.; Sajjad, M. Phân tích trình tự bộ gen COVID-19 để dự đoán và tạo ra biến thể mới. *Toán* 2022, 10, 4267. [\[CrossRef\]](#)

66. Quách, Y.; Chu, D.; Lý, W.; Cao, J.; Nie, R.; Xiong, L.; Ruan, X. Xác định tín hiệu polyadenylation bằng phương pháp nhúng sinh học thông qua mạng lưới đường cao tốc tích chập có cổng tự chú ý. ứng dụng. *Máy tính mềm.* 2021, 103, 107133. [\[Tham khảo chéo\]](#)

67. Vương, Y.; Hầu, Z.; Dương, Y.; Vương, Kc; Li, X. Nhận dạng và mô tả đặc điểm trên toàn bộ bộ gen của các chất tăng cường DNA với khung hợp nhất đa biến xếp chồng lên nhau. *Máy tính PLoS. Biol.* 2022, 18, e1010779. [\[Tham khảo chéo\]](#)

68. Năng, L.; Xu, K.; Hoàng, W.; Dương, YT; Lý, P.; Tăng, L.; Xiong, T.; Zhang, QC Dự đoán tương tác protein-RNA động của tế bào bằng cách học sâu bằng cách sử dụng cấu trúc RNA in vivo. *Độ phân giải tế bào.* 2021, 31, 495–516. [\[Tham khảo chéo\]](#)

69. Trương, TH; Hasib, MM; Chiu, YC; Hàn, ZF; Jin, YF; Flores, M.; Chen, Y.; Huang, Y. Transformer for Gene Expression Modeling (T-GEM): Mô hình học sâu có thể giải thích được để dự đoán kiểu hình dựa trên biểu hiện gen. *Ung thư* 2022, 14, 4763. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

70. Lee, D.; Dương, J.; Kim, S. Học mã histone với cửa sổ gen lớn và tương tác nhiễu sắc thể ba chiều bằng máy biến áp. *Nat. Cộng đồng.* 2022, 13, 6678. [\[Tham khảo chéo\]](#)

71. Chen, Y.; Xie, M.; Wen, J. Dự đoán biểu hiện gen từ sửa đổi histone với mạng lưới thần kinh dựa trên sự chú ý và học tập chuyển giao. *Đang trước. Genet.* 2022, 13, 1081842. [\[Tham khảo chéo\]](#)

72. Kang, M.; Lee, S.; Lee, D.; Kim, S. Học các cơ chế điều hòa gen dành riêng cho loại tế bào bằng phương pháp sâu dựa trên nhiều sự chú ý Học với không gian tiềm ẩn quy định. *Đang trước. Genet.* 2020, 11, 869. [\[Tham khảo chéo\]](#)

73. Liao, Y.; Quách, H.; Kinh, R.; Lạc, J.; Lý, M.; Li, Y. Dự đoán mức độ biểu hiện gen từ hồ sơ sửa đổi histone bằng phương pháp lai mạng học sâu. *Hóa chất. Trí tuệ. Phòng thí nghiệm. Hệ thống.* 2021, 219, 104456. [\[Tham khảo chéo\]](#)

74. Angenent-Mari, NM; Garruss, NHU; Soenksen, LR; Nhà thờ, G.; Collins, JJ Một phương pháp học sâu để chuyển mạch RNA có thể lập trình. *Nat. Cộng đồng.* 2020, 11, 5057. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

75. Tả, Z.; Vương, P.; Chen, X.; Thiên, L.; Ge, H.; Qian, D. SWnet: Một mô hình học sâu để dự đoán phản ứng thuốc từ dấu hiệu gen ung thư và cấu trúc hóa học phức hợp. *Thông tin sinh học BMC.* 2021, 22, 1–16. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

76. Karbalayghareh, A.; Sahin, M.; Leslie, CS Chromatin mô hình hóa quy định gen nhận biết tương tác với các mạng chú ý bằng đồ thị. *Bộ gen Res.* 2022, 32, 930–944. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

77. Phạm, TH; Thu, Y.; Tăng, J.; Xie, L.; Zhang, P. Một khung học sâu để sàng lọc hợp chất kiểu hình dựa trên cơ chế thông lượng cao và ứng dụng của nó vào việc tái sử dụng thuốc COVID-19. *Nat. Mach. Trí tuệ.* 2021, 3, 247–257. [\[Tham khảo chéo\]](#)

78. Dominic, N.; Cenggoro, TW; Budiarto, A.; Pardamean, B. Mạng lưới thần kinh đa gen sâu để dự đoán và xác định các gen liên quan đến năng suất ở các giống lúa Indonesia. *Khoa học. Dân biểu* 2022, 12, 13823. [\[CrossRef\]](#)

79. Lee, H.; Yeom, S.; Kim, S. BP-GAN: Dự đoán điểm nhánh của con người có thể giải thích được bằng cách sử dụng Mạng đối thủ sáng tạo có chú ý. *Truy cập IEEE* 2020, 8, 97851–97862. [\[Tham khảo chéo\]](#)

80. Lý, Z.; Chung, T.; Hoàng, D.; Bạ, ZH; Nie, R. Mạng chú ý biểu đồ phân cấp để dự đoán mối liên hệ giữa miRNA và bệnh. *Mol. Đó.* 2022, 30, 1775–1786. [\[Tham khảo chéo\]](#)

81. Bù, Y.; Giả, C.; Quách, X.; Lý, F.; Song, J. ĐỒNG: Một phương pháp học sâu tổng hợp để xác định các nguồn gốc vi-rút độc quyền các RNA can thiệp nhỏ ở thực vật. *Chức năng tóm tắt. Bộ gen.* 2023, 22, 274–280. [\[Tham khảo chéo\]](#)

82. Schapke, J.; Tavares, A.; Recamonde-Mendoza, M. EPGAT: Dự đoán tính thiết yếu của gen với mạng chú ý đồ thị. *Chuyển đổi IEEE-ACM. Máy tính. Biol. Thông tin sinh học.* 2022, 19, 1615–1626. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

83. Lưu, Y.; Yu, Y.; Zhao, S. Cơ chế chú ý kép và Phương pháp dựa trên mạng kết hợp tính năng để dự đoán bệnh LncRNA Hiệp hội. *Liên ngành-Khoa học-Máy tính. Khoa học cuộc sống.* 2022, 14, 358–371. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

84. Bài hát, J.; Thiên, S.; Yu, L.; Dương, Q.; Đại, Q.; Vương, Y.; Ngô, W.; Duan, X. RLF-LPI: Một khung học tập tổng hợp sử dụng thông tin trình tự để dự đoán tương tác lncRNA-protein dựa trên AE-ResLSTM và quyết định mờ. *Toán học. Sinh học. Anh.* 2022, 19, 4749–4764. [\[Tham khảo chéo\]](#)

85. Wekesa, JS; Mạnh, J.; Luân, Y. Một mô hình học sâu để dự đoán tương tác lncRNA-protein thực vật với sự chú ý của biểu đồ. *Mol. Genet. Bộ gen.* 2020, 295, 1091–1102. [\[Tham khảo chéo\]](#)

86. Ngô, H.; Pan, X.; Dương, Y.; Shen, HB Nhận biết vị trí gắn kết của các protein gắn với RNA có đặc tính kém trên RNA tròn chú ý sử dụng mạng Xiêm. *Ngắn gọn. Thông tin sinh học.* 2021, 22, bbab279. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

87. Dương, TH; Shiue, SC; Chen, KY; Tăng, YY; Wu, WS Xác định các mục tiêu piRNA trên mRNA ở C. Elegans bằng cách sử dụng kỹ thuật sâu mạng lưới chú ý nhiều đầu. *Thông tin sinh học BMC.* 2021, 22, 1–23. [\[Tham khảo chéo\]](#)

88. Lưu, T.; Zou, B.; Anh ấy, M.; Hu, Y.; Đậu, Y.; Thôi, T.; Tân, P.; Lý, S.; Rao, S.; Hoàng, Y.; et al. LncReader: Xác định các RNA không mã hóa dài có chức năng kép bằng cơ chế tự chú ý nhiều đầu. *Ngắn gọn. Thông tin sinh học.* 2022, 24, bbac579. [\[Tham khảo chéo\]](#)

89. Cao, M.; Shang, X. Xác định mối liên quan giữa lncRNA và tình trạng kháng thuốc dựa trên sự chú ý và học tập sâu cơ chế. *Đăng trước. Vi sinh vật.* 2023, 14, 1147778. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

90. Nguyễn, L.; Yang, Y. DeCban: Dự đoán các trang web tương tác CircRNA-RBP bằng cách sử dụng Mạng nhúng kép và Mạng chú ý liên nhánh. *Đăng trước. Genet.* 2021, 11, 632861. [\[Tham khảo chéo\]](#)

91. Bài hát, Z.; Hoàng, D.; Bài hát, B.; Chen, K.; Song, Y.; Lưu, G.; Su, J.; de Magalhaes, JP; Rigden, DJ; Meng, J. Mạng lưới thần kinh đa nhân dựa trên sự chú ý để dự đoán và giải thích tích hợp về mười hai sửa đổi RNA xảy ra rộng rãi. *Nat. Cộng đồng.* 2021, 12, 4011. [\[Tham khảo chéo\]](#)

92. Chen, K.; Chu, X.; Vương, J.; Hào, L.; Lưu, Z.; Liu, Y. ncDENSE: Một phương pháp tính toán mới dựa trên khung học sâu để dự đoán họ RNA không mã hóa. *Thông tin sinh học BMC.* 2023, 24, 1–20. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

93. Dương, Y.; Hầu, Z.; Ma, Z.; Lý, X.; Wong, KC iCircRBP-DHN: Xác định các vị trí tương tác CircRNA-RBP bằng cách sử dụng deep mạng phân cấp. *Ngắn gọn. Thông tin sinh học.* 2021, 22, bbac274. [\[Tham khảo chéo\]](#)

94. Lý, G.; Vương, D.; Trương, Y.; Lương, C.; Tiểu, Q.; Luo, J. Sử dụng Mạng chú ý đồ thị và Mạng tích chập đồ thị để khám phá mối liên hệ giữa bệnh CircRNA và bệnh tật ở người dựa trên dữ liệu đa nguồn. *Đăng trước. Genet.* 2022, 13, 829937. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

95. Vương, H.; Hàn, J.; Lý, H.; Duẩn, L.; Lưu, Z.; Cheng, H. CDA-SKAG: Dự đoán mối liên hệ giữa bệnh tuần hoàn-RNA bằng cách sử dụng phản ứng tổng hợp hạt nhân tương tự và bộ mã hóa biểu đồ tự động tăng cường sự chú ý. *Toán học. Sinh học. Anh.* 2023, 20, 7957–7980. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

96. Lý, G.; Lin, Y.; Lạc, J.; Tiểu, Q.; Liang, C. GGAECDA: Dự đoán mối liên hệ giữa bệnh tuần hoàn-RNA bằng cách sử dụng bộ mã hóa tự động biểu đồ dựa trên việc học biểu diễn biểu đồ. *Máy tính. Biol. Chem.* 2022, 99, 107722. [\[Tham khảo chéo\]](#)

97. Fan, Y.; Chen, M.; Pan, X. GCRFLDA: Ghi điểm liên quan đến bệnh lncRNA bằng cách sử dụng hoàn thành ma trận tích chập biểu đồ với trường ngẫu nhiên có điều kiện. *Ngắn gọn. Thông tin sinh học.* 2022, 23, bbab361. [\[Tham khảo chéo\]](#)

98. Thịnh, N.; Thôi, H.; Trương, T.; Xuân, P. Mã hóa biểu diễn đa cấp chú ý dựa trên tích chập và phương sai bộ mã hóa tự động để dự đoán mối liên quan giữa lncRNA và bệnh. *Ngắn gọn. Thông tin sinh học.* 2021, 22, bbab067. [\[Tham khảo chéo\]](#)

99. Ngưu, M.; Zou, Q.; Lin, C. CRBPDl: Xác định các vị trí tương tác CircRNA-RBP bằng cách sử dụng phương pháp mạng lưới thần kinh tổng hợp. *Máy tính PLoS. Biol.* 2022, 18, e1009798. [\[Tham khảo chéo\]](#)

100. Trương, Y.; Bạ, F.; Gao, X. MCA-Net: Mạng thần kinh tích chập chú ý và mã hóa đa tính năng để dự đoán Hiệp hội bệnh lncRNA. *Chuyển đổi IEEE-ACM. Máy tính. Biol. Thông tin sinh học.* 2022, 19, 2907–2919. [\[Tham khảo chéo\]](#)

101. Lưu, Y.; Fu, Q.; Bành, X.; Chu, C.; Lưu, G.; Liu, L. Học tập đa trường sâu dựa trên sự chú ý để phân loại RNA tròn và RNA không mã hóa dài khác. *Gen* 2021, 12, 2018. [\[Tham khảo chéo\]](#)

102. Quách, Y.; Lỗi, X.; Lưu, L.; Pan, Y. Circ2CBA: Dự đoán các vị trí liên kết CircRNA-RBP kết hợp cơ chế học tập sâu và chú ý. *Đăng trước. Máy tính. Khoa học.* 2023, 17, 175904. [\[Tham khảo chéo\]](#)

103. Clauwaert, J.; Menschaert, G.; Waegeman, W. Khả năng giải thích trong các mô hình biến áp cho bộ gen chức năng. *Ngắn gọn. Thông tin sinh học.* 2021, 22, bbab060. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

104. Phùng, P.; Tiêu, A.; Phương, M.; Vạn, F.; Lý, S.; Lang, P.; Triệu, D.; Zeng, J. Một khung dựa trên máy học để lập mô hình sự kéo dài phiên mã. *Proc. Natl. Học viện. Khoa học. Hoa Kỳ* 2021, 118, e2007450118. [\[Tham khảo chéo\]](#)

105. Hàn, K.; Thần, LC; Chu, YH; Xu, J.; Bài hát, J.; Yu, DJ MARENet: Dự đoán các trang web liên kết yếu tổ phiên mã bằng cách kết hợp sự chú ý từ dưới lên và từ trên xuống theo nhiều quy mô và mạng dự. *Ngắn gọn. Thông tin sinh học.* 2022, 23, bbab445. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

106. Tao, Y.; Mã, X.; Palmer, D.; Schwartz, R.; Lữ, X.; Osmanbeyoglu, HU Học sâu có thể giải thích để suy luận dựa trên chất nhuộm sắc về các chương trình phiên mã được thúc đẩy bởi sự thay đổi soma ở các bệnh ung thư. *Axit nucleic Res.* 2022, 50, 10869–10881. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

107. Asim, MN; Ibrahim, MA; Malik, MI; Zehe, C.; Cloarec, O.; Cổ gắng, J.; Dengel, A.; Ahmed, S. EL-RMLocNet: Mạng LSTM có thể giải thích được để dự đoán nội địa hóa nhiều ngăn liên quan đến RNA. *Máy tính. Cấu trúc. Công nghệ sinh học. J.* 2022, 20, 3986–4002. [\[Tham khảo chéo\]](#)

108. Công viên, S.; Koh, Y.; Jeon, H.; Kim, H.; Yeo, Y.; Kang, J. Nâng cao khả năng diễn giải của dự đoán trang web liên kết yếu tổ phiên mã sử dụng cơ chế chú ý. *Khoa học. Dân biểu* 2020, 10, 13413. [\[CrossRef\]](#)

109. Yển, Z.; Lecuyer, E.; Blanchette, M. Dự đoán nội địa hóa mRNA dưới tế bào bằng cách sử dụng mạng lưới thần kinh tái phát sâu. *Tin sinh học* 2019, 35, I333–I342. [\[Tham khảo chéo\]](#)

110. Bài hát, J.; Thiên, S.; Yu, L.; Xing, Y.; Dương, Q.; Duẩn, X.; Dai, Q. AC-Caps: Mạng lưới viên nang dựa trên sự chú ý để dự đoán RBP Trang web liên kết của lncRNA. *Liên ngành. Khoa học. Máy tính. Khoa học cuộc sống.* 2020, 12, 414–423. [\[Tham khảo chéo\]](#)

111. Công, P.; Trình, L.; Trương, Z.; Mạnh, A.; Nói dối.; Chen, J.; Zhang, L. Phương pháp tích hợp đa omics dựa trên mạng học sâu chú ý để phân loại dữ liệu y sinh. *Máy tính. Phương pháp Chương trình. Sinh học.* 2023, 231, 107377. [\[Tham khảo chéo\]](#)

112. Kayikci, S.; Khoshgoftaar, TM Dự đoán ung thư vú bằng cách sử dụng phương pháp học sâu đa phương thức có kiểm soát. *J. Dữ liệu lớn* 2023, 10, 1–11. [\[Tham khảo chéo\]](#)

113. Anh, L.; Trương, Y.; Dương, X.; Thần, F.; Xu, B. Phương pháp dự đoán gen nhạy cảm với ung thư buồng trứng dựa trên học sâu Phương pháp. *Đăng trước. Nhà phát triển di động Biol.* 2021, 9, 730475. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

114. Kang, Q.; Mạnh, J.; Shi, W.; Luan, Y. Ensemble Deep Learning dựa trên việc nâng cao thông tin đa cấp và quyết định mờ tham lam để dự đoán tương tác miRNA-lncRNA thực vật. *Liên ngành. Khoa học. Máy tính. Khoa học cuộc sống.* 2021, 13, 603–614. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

115. Vương, C.; Lye, X.; Kaalia, R.; Kumar, P.; Rajapakse, JC Phương pháp học sâu và đa omics để dự đoán phản ứng của thuốc trong bệnh ung thư. Thông tin sinh học BMC. 2022, 22, 1–15. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

116. Chấn, YH; Vương, C.; Soh, WK; Rajapakse, JC Kết hợp bộ dữ liệu hình ảnh thần kinh và Omics để phân loại bệnh bằng cách sử dụng Mạng lưới thần kinh đồ thị. Đăng trước. Khoa học thần kinh. 2022, 16, 866666. [\[Tham khảo chéo\]](#)

117. Lưu, Q.; Anh ấy, D.; Xie, L. Dự đoán tính đặc hiệu ngoài mục tiêu và mức độ phù hợp dành riêng cho tế bào của Hệ thống CRISPR-Cas bằng cách tăng cường sự chú ý học sâu và tính năng gen dựa trên mạng. Máy tính PLoS. Biol. 2019, 15, e1007480. [\[Tham khảo chéo\]](#)

118. Lưu, X.; Vương, S.; Ai, D. Dự đoán kết quả sửa chữa CRISPR/Cas9 bằng Khung học tập sâu dựa trên sự chú ý. Tế bào 2022, Ngày 11 tháng 1 năm 1847. [\[Tham khảo chéo\]](#)

119. Vạn, Y.; Jiang, Z. TransCrispr: Mô hình lai dựa trên máy biến áp để dự đoán sự phân tách RNA hướng dẫn đơn CRISPR/Cas9 Hiệu quả. Chuyển đổi IEEE-ACM. Máy tính. Biol. Thông tin sinh học. 2023, 20, 1518–1528. [\[Tham khảo chéo\]](#)

120. Tiểu, LM; Vạn, YQ; Jiang, ZR AttCRISPR: Một mô hình có thể diễn giải theo không thời gian để dự đoán hoạt động nhắm mục tiêu của sgRNA. Thông tin sinh học BMC . 2021, 22, 1–17. [\[Tham khảo chéo\]](#)

121. Mathis, N.; Allam, A.; Kissling, L.; Marquart, KF; Schmidheini, L.; Solari, C.; Balazs, Z.; Krauthammer, M.; Schwank, G. Dự đoán hiệu quả chỉnh sửa chính và độ tinh khiết của sản phẩm bằng cách học sâu. Nat. Công nghệ sinh học. 2023. [\[Tham khảo chéo\]](#)

122. Trương, ZR; Jiang, ZR Sử dụng hiệu quả thông tin trình tự để dự đoán CRISPR-Cas9 ngoài mục tiêu. Máy tính. Cấu trúc. Công nghệ sinh học. J. 2022, 20, 650–661. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

123. Trương, G.; Tăng, T.; Đại, Z.; Dai, X. Dự đoán về tính đặc hiệu và hiệu quả phân tách RNA dẫn hướng đơn CRISPR/Cas9 bằng các mạng lưới thần kinh tích chập dựa trên sự chú ý. Máy tính. Cấu trúc. Công nghệ sinh học. J. 2021, 19, 1445–1457. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

124. Lin, Z.; Ou-Yang, L. Suy ra mạng lưới điều hòa gen từ dữ liệu biểu hiện gen đơn bào thông qua phép tương phản đa góc nhìn sâu học hỏi. Ngắn gọn. Thông tin sinh học. 2022, 24, bbac586. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

125. Xu, J.; Trương, A.; Lưu, F.; Zhang, X. STGRNS: Một phương pháp dựa trên máy biến áp có thể giải thích được để suy ra mạng lưới điều hòa gen từ dữ liệu phiên mã đơn bào. Tin sinh học 2023, 39, btad165. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

126. Phùng, X.; Fang, F.; Long, H.; Tăng, R.; Yao, Y. Phân tích dữ liệu RNA-seq đơn bào bằng cách sử dụng bộ mã hóa tự động đồ thị và mạng chú ý đồ thị. Đăng trước. Genet. 2022, 13, 1003711. [\[Tham khảo chéo\]](#)

127. Ullah, F.; Ben-Hur, A. Một mô hình tự chú ý để suy ra sự hợp tác giữa các tính năng quản lý. Axit nucleic Res. 2021, 49, e77. [\[Tham khảo chéo\]](#)

128. Tạ, X.; Vương, Y.; Thịnh, N.; Trương, S.; Cao, Y.; Fu, Y. Dự đoán mối liên quan giữa miRNA và bệnh dựa trên thông tin nhiều góc nhìn dung hợp. Đăng trước. Genet. 2022, 13, 979815. [\[Tham khảo chéo\]](#)

129. Lee, M. Một mô hình học tập sâu tập hợp với cơ chế chú ý gen để ước tính tiên lượng của học sinh cấp thấp Bệnh u thần kinh đệm. Sinh học 2022, 11, 586. [\[Tham khảo chéo\]](#)

130. Choi, SR; Lee, M. Ước tính Tiên lượng của U thần kinh đệm cấp độ thấp với sự chú ý đến gen bằng cách sử dụng Multi-Omics và Multi-Modal Đề án. Sinh học 2022, 11, 1462. [\[Tham khảo chéo\]](#)

131. Dutta, P.; Patra, AP; Saha, S. DeepPROG: Mô hình dựa trên sự chú ý sâu sắc để tiên lượng gen bệnh bằng cách kết hợp dữ liệu đa Omics. Chuyển đổi IEEE-ACM. Máy tính. Biol. Thông tin sinh học. 2022, 19, 2770–2781. [\[Tham khảo chéo\]](#)

132. Xing, X.; Dương, F.; Lý, H.; Trương, J.; Triệu, Y.; Cao, M.; Hoàng, J.; Yao, J. Mạng lưới thần kinh biểu đồ chú ý đa cấp dựa trên các mô-đun gen đồng biểu hiện để chẩn đoán và tiên lượng bệnh. Tin sinh học 2022, 38, 2178–2186. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

133. Mạnh, X.; Vương, X.; Trương, X.; Trương, C.; Trương, Z.; Trương, K.; Wang, S. Một mô hình sinh tồn Cox dựa trên cơ chế chú ý mới bằng cách khai thác thông tin gen thực nghiệm về bệnh ung thư. Tế bào 2022, 11, 1421. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

134. Phùng, C.; Tường, T.; Yi, Z.; Mạnh, X.; Chu, X.; Hoàng, G.; Triệu, X.; Chen, F.; Xiong, B.; Feng, J. Một mô hình học sâu với cơ chế chú ý có thể dự đoán chính xác tỷ lệ sống sót trong bệnh u nguyên bào thần kinh. Đăng trước. Oncol. 2021, 11, 653863. [\[Tham khảo chéo\]](#)

135. Gokhale, M.; Mohanty, SK; Ojha, A. GeneViT: Máy biến đổi thị giác gen với DeepInsight cải tiến để phân loại ung thư. Máy tính. Biol. Med. 2023, 155, 106643. [\[Tham khảo chéo\]](#)

136. Beykikhoshk, A.; Quinn, TP; Lee, SC; Trần, T.; Venkatesh, S. DeepTRIAGE: Điểm số dấu ấn sinh học có thể giải thích và cá nhân hóa bằng cách sử dụng cơ chế chú ý để phân loại các phân nhóm ung thư vú. BMC Med. Bộ gen. 2020, 13, 1–10. [\[Tham khảo chéo\]](#)

137. Manica, M.; Oskoei, A.; Sinh ra, J.; Subramanian, V.; Saez-Rodriguez, J.; Martinez, MR Hướng tới dự đoán độ nhạy của hợp chất chống ung thư có thể giải thích được thông qua Bộ mã hóa tích chập dựa trên sự chú ý đa phương thức. Mol. Dược phẩm. 2019, 16, 4797–4806. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

138. Lee, S.; Lim, S.; Lee, T.; Sung, tôi.; Kim, S. Phân loại và mô hình hóa phân nhóm ung thư bằng cách chú ý và lan truyền con đường. Tin sinh học 2020, 36, 3818–3824. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

139. Hầu, Z.; Dương, Y.; Lý, H.; Vương, KC; Li, X. iDeepSubMito: Xác định quá trình nội địa hóa protein trong màng sụn bằng phương pháp học sâu. Ngắn gọn. Thông tin sinh học. 2021, 22, bbab288. [\[Tham khảo chéo\]](#)

140. Công, H.; Ôn, J.; Lạc, R.; Phong, Y.; Quách, J.; Phú, H.; Chu, X. Tích hợp tối ưu hóa trình tự mRNA bằng cách sử dụng học sâu. Ngắn gọn. Thông tin sinh học. 2023, 24, bbad001. [\[Tham khảo chéo\]](#)

141. Armenteros, JJA; Salvatore, M.; Emanuelsson, O.; Winther, O.; von Heijne, G.; Elofsson, A.; Nielsen, H. Phát hiện trình tự tín hiệu trong việc nhắm mục tiêu peptide bằng cách sử dụng học sâu. Khoa học cuộc sống. Liên minh 2019, 2, e201900429. [\[Tham khảo chéo\]](#)

142. Littmann, M.; Heinzinger, M.; Dallago, C.; Weissnow, K.; Rost, B. Việc nhúng protein và học sâu dự đoán dự lượng liên kết cho các loại phối tử khác nhau. Khoa học. Dân biểu 2021, 11, 23916. [\[CrossRef\]](#) [\[PubMed\]](#)

143. Song, T.; Đại, H.; Vương, S.; Vương, G.; Trương, X.; Trương, Y.; Jiao, L. TransCluster: Phương pháp nhận dạng loại tế bào cho dữ liệu RNA-Seq đơn bào sử dụng phương pháp học sâu dựa trên máy biến áp. Đăng trước. Genet. 2022, 13, 1038919. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)

144. Phùng, X.; Trương, H.; Lâm, H.; Phân tích dữ liệu RNA-seq đơn bào dài, H. dựa trên mạng lưới thần kinh đồ thị có hướng. *Phương pháp* 2023, 211, 48–60. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)
145. Buterez, D.; Bica, tôi.; Tariq, tôi.; Andres-Terre, H.; Lio, P. CellVGAE: Quy trình phân tích scRNA-seq không giám sát với biểu đồ mạng lưới chú ý. *Tin sinh học* 2022, 38, 1277–1286. [\[Tham khảo chéo\]](#)
146. Trương, Y.; Blanchette, M. Chú thích cấu trúc topo hướng dẫn bằng tham chiếu của dữ liệu Hi-C. *Nat. Cộng đồng*. 2022, 13, 7426. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)
147. Schwarz, K.; Allam, A.; Gonzalez, ngừ trúa; Krauthammer, M. Chú ýDDI: Phương pháp học sâu dựa trên sự chú ý của Siam dành cho dự đoán tương tác thuốc-thuốc. *Thông tin sinh học BMC*. 2021, 22, 1–19. [\[Tham khảo chéo\]](#)
148. Kim, E.; Nam, H. DeSIDE-DDI: Dự đoán có thể giải thích được về tương tác thuốc-thuốc bằng cách sử dụng biểu hiện gen do thuốc gây ra. *J. Cheminform*. 2022, 14, 1–12. [\[Tham khảo chéo\]](#)
149. Lưu, Q.; Xie, L. TranSynergy: Mạng lưới thần kinh sâu có thể giải thích được điều khiển bằng cơ chế để dự đoán và lộ trình tổng hợp giải mã sự kết hợp thuốc. *Máy tính PLoS. Biol.* 2021, 17, e1008653. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)
150. Vương, J.; Lưu, X.; Thần, S.; Đặng, L.; Liu, H. DeepDDS: Mạng lưới thần kinh đồ thị sâu với cơ chế chú ý để dự đoán sự phối hợp phối hợp thuốc. *Ngắn gọn. Thông tin sinh học*. 2022, 23, bbab390. [\[Tham khảo chéo\]](#)
151. Yu, G.; Tăng, J.; Vương, J.; Trương, H.; Trương, X.; Guo, M. Mất cân bằng học tập đa phiên bản sâu để dự đoán các tương tác đồng dạng-dồng dạng. *Int. J. Trí tuệ. Hệ thống*. 2021, 36, 2797–2824. [\[Tham khảo chéo\]](#)
152. Yamaguchi, H.; Các giao thức Saito, Y. Evotuning để dự đoán hiệu ứng biến thể dựa trên Máy biến áp trên các protein đa miền. *Ngắn gọn. Thông tin sinh học*. 2021, 22, bbab234. [\[Tham khảo chéo\]](#)
153. Chu, J.; Chen, Q.; Braun, P.; Mandell, KAP; Jaffe, AE; Tân, HY; Hyde, TM; Kleinman, JE; Kali, JB; Shinozaki, G.; et al. Học sâu dự đoán các biến thể điều hòa quá trình methyl hóa DNA trong não người và làm sáng tỏ tính di truyền của các rối loạn tâm thần. *Proc. Natl. Học viện. Khoa học. Hoa Kỳ* 2022, 119, e2206069119. [\[Tham khảo chéo\]](#)
154. Cao, L.; Trương, Q.; Song, H.; Lin, K.; Pang, E. DeepASmRNA: Dự đoán không cần tham chiếu về các sự kiện ghép nối thay thế với mô hình học sâu có thể mở rộng và có thể hiểu được. *iScience* 2022, 25, 105345. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)
155. Gupta, S.; Shankar, R. miWords: Học sâu tổng hợp dựa trên máy biến áp để khám phá có độ chính xác cao về các vùng tiền miRNA trên bộ gen thực vật. *Ngắn gọn. Thông tin sinh học*. 2023, 24, bbad088. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)
156. Trương, ZY; Ninh, L.; Bạ, X.; Dương, YH; Futamura, Y.; Sakurai, T.; Lin, H. iLoc-miRNA: Dự đoán miRNA ngoại bào/nội bào bằng cách sử dụng BiLSTM sâu với cơ chế chú ý. *Ngắn gọn. Thông tin sinh học*. 2022, 23, bbac395. [\[Tham khảo chéo\]](#) [\[PubMed\]](#)
157. Choi, JM; Chae, H. moBRCA-net: Khung phân loại phân nhóm ung thư vú dựa trên mạng lưới thần kinh chú ý đa omics. *Thông tin sinh học BMC*. 2023, 24, 1–15. [\[Tham khảo chéo\]](#)
158. Âm, C.; Chen, Z. Phát triển phân loại bệnh bền vững thông qua học sâu và học bán giám sát. *Chăm sóc sức khỏe* 2020, 8, 291. [\[CrossRef\]](#)
159. Song, H.; Âm, C.; Lý, Z.; Phong, K.; Cao, Y.; Chàng trai.; Sun, H. Xác định các gen điều khiển ung thư bằng cách tích hợp dữ liệu đa hệ với Mạng thần kinh đồ thị. *Chất chuyển hóa* 2023, 13, 339. [\[CrossRef\]](#)
160. Bài hát, JT; Woo, DU; Lee, Y.; Choi, SH; Kang, YJ Chiến lược học máy bán giám sát trên họ gen Sự đa dạng để làm sáng tỏ sự tổng hợp Resveratrol. *Thực vật* 2021, 10, 2058. [\[Tham khảo chéo\]](#)
161. Munoz, SA; Park, J.; Stewart, CM; Martin, sáng; Hedengren, JD Học chuyển sâu để dự đoán mô hình gần đúng Điều khiển. *Quy trình* 2023, 11, 197. [\[CrossRef\]](#)
162. Dastour, H.; Hassan, QK So sánh các phương pháp học chuyển giao sâu để phân loại sử dụng đất và che phủ đất. *Tính bền vững* 2023, 15, 7854. [\[Tham khảo chéo\]](#)
163. Dương, L.; Hoàng, R.; Trương, J.; Hoàng, J.; Vương, L.; Đồng, J.; Shao, J. Chuyển giao phương pháp học sâu được đào tạo trước xuyên lục địa Mô hình lập bản đồ lúa gạo và khả năng khái quát hóa của nó. *Cảm biến từ xa* 2023, 15, 2443. [\[Tham khảo chéo\]](#)

Tuyên bố miễn trừ trách nhiệm/Lưu ý của nhà xuất bản: Các tuyên bố, ý kiến và dữ liệu trong tất cả các ấn phẩm chỉ là của (các) tác giả và (các) cộng tác viên chứ không phải của MDPI và/hoặc (các) biên tập viên. MDPI và/hoặc (các) biên tập viên từ chối trách nhiệm đối với mọi thương tích đối với con người hoặc tài sản do bất kỳ ý tưởng, phương pháp, hướng dẫn hoặc sản phẩm nào được đề cập trong nội dung.