

Bài báo khoa học

Phân tích chuỗi dữ liệu nghiệm triều sử dụng mạng nơ ron hồi tiếp với nút có cổng (GRU)

Nguyễn Gia Trọng^{1,2*}, Bùi Ngọc Quý³

¹ Trường Đại học Mở - Địa chất; nguyengiatrong@humg.edu.vn

² Nhóm nghiên cứu Trắc địa cao cấp - môi trường

³ Viện Tài nguyên và Môi trường, Đại học Quốc gia Hà Nội; quybncre@vnu.edu.vn

*Tác giả liên hệ: nguyengiatrong@humg.edu.vn; Tel.: +84-963124980

Ban Biên tập nhận bài: 15/3/2024; Ngày phản biện xong: 27/4/2024; Ngày đăng bài: 25/9/2024

Tóm tắt: Nghiên cứu này đã ứng dụng mô hình GRU (*Gated Recurrent Unit*) thuộc nhóm mạng nơ ron hồi quy (*RNN - Recurrent Neural Network*) để phân tích dữ liệu nghiệm triều thu nhận được tại trạm nghiệm triều Vũng Tàu trong thời gian từ 01/01/1999 đến 31/12/2022. Kết quả thực nghiệm cho thấy, các đặc trưng thống kê về hiệu suất của mô hình đạt kết quả rất tốt: RMSE = 2,2 mm; MAE = 0,5 mm; Kappa = 0,98 và F-Score = 0,96. Bên cạnh đó, để khẳng định liệu mô hình có thực sự hiệu quả hay không đã phân chia bộ dữ liệu thành tập dữ liệu huấn luyện, tập dữ liệu kiểm tra và tập dữ liệu dự báo. Kết quả thống kê cho thấy mô hình GRU có thể dự đoán một cách tin cậy giá trị triều thông qua các đặc trưng thống kê của tập dữ liệu dự báo như RMSE = 0,06 mm, MAE = 0,05 mm. Nghiên cứu cũng chỉ ra rằng, cần thiết phải thực thi phát hiện và loại trừ ngoại lai đối với tập hợp dữ liệu đầu vào để có thể thu được kết quả dự báo đạt độ chính xác cao.

Từ khóa: RNN; GRU; Dữ liệu nghiệm triều; Chuỗi dữ liệu theo thời gian; AI.

1. Mở đầu

Hiện tượng ấm lên toàn cầu là yếu tố chính gây ra sự dâng cao của mực nước biển, được thúc đẩy bởi việc thải ra lượng khí nhà kính ngày càng tăng từ các hoạt động của con người như đốt nhiên liệu hóa thạch và phá rừng. Hiện tượng này làm cho các phân tử nước biển mở rộng khi chúng hấp thụ nhiệt độ khí quyển, đóng góp đáng kể vào việc nâng cao mực nước biển. Đồng thời, sự tan chảy của băng và tuyết, đặc biệt là tại các vùng cực như Greenland và Nam Cực, tiếp tục gia tăng, giải phóng ra lượng lớn nước ngọt vào biển và đại dương. Ngoài ra, sự mất mát của lớp băng trong đất, bao gồm các tảng băng và băng chài, cũng tăng do nhiệt độ tăng lên, làm tăng thêm sự nâng cao của mực nước biển. Hơn nữa, các thay đổi trong sử dụng đất, như khai thác nước dưới đất và xây dựng đập, gây ra sự sụt giảm đất đai, làm gia tăng hiện tượng sụt lún ven biển và làm tăng sự xâm nhập của biển [1].

Sự dâng lên của mực nước biển đe dọa các vùng ven biển có địa hình thấp và các hòn đảo gây ra thiệt hại tài sản nghiêm trọng, mất mát môi trường sống và di dời cộng đồng. cũng đặt ra mối đe dọa lớn khi mực nước biển tăng cao xâm nhập vào các nguồn nước ngọt, đe dọa cung cấp nước sạch và tổ chức cấu trúc sinh thái. Làm thay đổi môi trường ven biển, bao gồm các đầm lầy, rừng ngập mặn và rạn san hô, đe dọa đa dạng sinh học biển và các hệ sinh thái quan trọng ... [2]. Bên cạnh ý nghĩa trong nghiên cứu tại biển thiên tai, xác định được giá trị tăng lên của mực nước biển cũng góp phần cập nhật các bề mặt động lực đặc trưng trên biển. Do đó, quan trắc mực nước biển, bao gồm sự thay đổi của mực nước biển là công việc được tiến hành thường xuyên và liên tục. Có nhiều công nghệ khác nhau được sử dụng để quan trắc sự thay đổi của mực nước biển bao gồm sử dụng trạm nghiệm triều, dữ liệu đo

cao vệ tinh, GNSS phản xạ,...[3]. Từ năm 2003, máy đo thủy triều kỹ thuật số với nhiều sensor khác nhau đã được áp dụng để nâng cao độ chính xác và độ tin cậy của kết quả quan trắc. Chất lượng của bộ dữ liệu và tính sẵn có trực tuyến của nó mang lại khả năng mở rộng nghiên cứu sang các lĩnh vực nghiên cứu khác, chẳng hạn như xác định thời gian thực về nước dâng do bão và sóng thần. Chuỗi thời gian hoàn chỉnh, bao gồm cả dữ liệu tương tự và dữ liệu số, đã được tích hợp, xử lý và phân tích và đã thu được kết quả sơ bộ mới [4]. Bằng cách kết hợp chuỗi dữ liệu nghiệm triều, dữ liệu thu được từ máy đo độ cao SIRAL trên vệ tinh CryoSat-2 đã góp phần nâng cao độ chính xác đo độ cao mực nước ven biển tại Vương Quốc Anh lên mức 4 cm [5]. Các trạm nghiệm triều sử dụng trong nghiên cứu sự dâng lên của mực nước biển thường là các trạm quan trắc liên tục trong rất nhiều năm, có trạm lên tới hàng trăm năm. Do đó, dữ liệu nghiệm triều ở đây là dữ liệu theo chuỗi thời gian [6].

Để phân tích chuỗi dữ liệu nghiệm triều, các công cụ phân tích truyền thông và hiện đại (như trí tuệ nhân tạo) đã được sử dụng. Nghiên cứu [6] đã phân tích sự thay đổi của mực nước biển theo tháng, theo mùa và theo năm trên cơ sở đó đã xác định được sự thay đổi tương đối của mực nước biển tại khu vực Manila giai đoạn 1902-1962 là 1,6 mm/năm và tương ứng là 15,6 mm/năm cho giai đoạn 1963-1993. Các phương pháp phân tích giá trị đơn (SVD), ứng dụng phép lọc Kalman, kỹ thuật hàm điều hòa cũng đã được áp dụng để phân tích dữ liệu nghiệm triều [7].

Đã có nhiều mô hình trí tuệ nhân tạo khác nhau được ứng dụng để phân tích dữ liệu nghiệm triều. Nghiên cứu [8] đã phát triển mô hình dự báo triều ngắn hạn dựa trên sự kết hợp giữa mô hình trí tuệ nhân tạo ANN với mô hình thủy động lực học để dự báo ngắn hạn mực nước biển nhằm giảm thiểu ảnh hưởng của tai biến thiên tai ven biển.

Với các vùng đất thấp ven biển hoặc mục tiêu quản lý vùng ven biển thì cảnh báo mực nước hàng giờ là công việc rất quan trọng. Nghiên cứu [9] đã ứng dụng hai mô hình là ELM (*Extreme Learning Machine*) và RVM (*Relevance Vector Machine*) để dự đoán sự thay đổi của mực nước biển với sai số trung phương tối thiểu là 34,73 mm và $R^2 = 0,93$.

Với phương pháp tiếp cận khác, nghiên cứu [10] đã sử dụng dữ liệu đầu vào là hệ số điều hòa triều và điều kiện khí tượng thu nhận được bởi hơn 600 trạm nghiệm triều trên thế giới để dự đoán tổng mực nước biển. Kết quả cho thấy kỹ năng cao nhất quán trên toàn cầu của mạng lưới thần kinh (NN) trong việc nắm bắt sự biến đổi của biển tại các máy đo trên toàn cầu. Trong khi động lực chính do khí quyển điều khiển có thể được nắm bắt bằng các hồi quy tuyến tính đa biến, thì sự tăng cường do khí quyển, nước dâng và phi tuyến tính của thủy triều trong môi trường ven biển phức tạp chỉ được dự đoán bằng NN [10].

Mô hình LSTM đã được ứng dụng để phân tích dữ liệu nghiệm triều trong 21 năm tại 17 bến cảng của Đài Loan. Để đánh giá hiệu quả của mô hình sử dụng các đại lượng sai số phần trăm tuyệt đối trung bình (MAPE) và sai số bình phương trung bình gốc (RMSE) của kết quả dự báo. Kết quả cho thấy giá trị đạt được lần lượt là 6,97% và 0,049 m cho thấy mô hình có thể giảm thiểu một cách hiệu quả các vấn đề chông chéo do phương pháp học máy gây ra trong dự báo liên tục [11]. Nghiên cứu [12] đã khảo sát hiệu quả của việc đưa thêm dữ liệu khí tượng khi sử dụng các mô hình MRL, ANFIS trong dự đoán mực nước. Kết quả thực nghiệm cho thấy, khi đưa thêm các yếu tố khí tượng làm tham số đầu vào sẽ tăng độ chính xác hiệu suất của mô hình MLR lên tới 33% để dự đoán mực nước biển ngắn hạn.

Áp dụng phương pháp Phân tích chế độ theo kinh nghiệm của Ensemble (EEMD) kết hợp với mô hình học máy LSTM cho kết quả dự đoán mực nước vượt trội so với sử dụng độc lập mô hình LSTM [13]. Nghiên cứu [14] đã chỉ ra rằng, sử dụng mô hình mạng thần kinh tự liên kết (AANN) cho phép phát hiện các hiện tượng cực đoan trên biển dựa trên việc quan trắc mực nước biển.

Việt Nam là quốc gia chịu ảnh hưởng trầm trọng do hiện tượng nước biển dâng do đó có nhiều nghiên cứu liên quan đã được triển khai. Nghiên cứu [15] đã phân tích mối tương quan giữa các điều kiện khí tượng với mực nước biển phục vụ đánh giá ảnh hưởng của biến đổi khí hậu tại khu vực. Nghiên cứu [16] trên cơ sở phân tích chuỗi dữ liệu vệ tinh từ năm 1993

đến năm 2010 thu nhận bởi Trung tâm dự báo môi trường quốc gia Hoa Kỳ đã làm sáng tỏ được phần nào mối tương quan giữa nhiệt độ và sự dâng lên của mực nước tại biển Đông. Nghiên cứu [17] cũng đã phân tích mối tương quan giữa điều kiện khí tượng với hiện tượng triều cường làm cơ sở để dự báo dị thường mực nước biển.

Trên cơ sở dữ liệu đo thủy triều trong 47 năm, ứng dụng phép lọc Kalman các tác giả [18] đã xác định được giá trị tăng của mực nước biển tại Hạ Long giai đoạn 1974-2020 là 3,14 mm/năm và tương ứng là 5,83 mm/năm cho giai đoạn từ năm 2002 đến năm 2020. Nghiên cứu [19] đã ứng dụng phiên bản cải tiến của mạng nơ ron hồi quy để dự báo độ cao sóng với hệ số tương quan lớn nhất là $R^2 = 0,582$.

Có thể thấy rằng, trí tuệ nhân tạo chưa được ứng dụng rộng rãi trong phân tích chuỗi dữ liệu nghiệm triều tại Việt Nam. Nghiên cứu này nghiên cứu ứng dụng mô hình mạng nơ ron hồi quy cải tiến (GRU) để phân tích dữ liệu nghiệm triều tại Việt Nam.

2. Dữ liệu và phương pháp nghiên cứu

2.1. Dữ liệu nghiên cứu

Dữ liệu dùng trong nghiên cứu này được cung cấp bởi Đài Khí tượng thủy văn Nam Bộ thành phố Hồ Chí Minh thu nhận được bởi trạm nghiệm triều tại Vũng Tàu (hình 1). Từ dữ liệu gốc ban đầu, được thu thập với giãn các 2 giờ 1 lần được tính toán quy chuyên về giá trị triều trung bình theo ngày. Cấu trúc dữ liệu lúc này được cho dưới dạng ngày, tháng, năm và giá trị độ cao mực nước. Để có thể phân tích sử dụng mô hình trí tuệ nhân tạo, cần chuyển các yếu tố thời gian về dạng ngày ở dạng số thực.

2.2 Phương pháp nghiên cứu

Quy trình tính toán trong nghiên cứu này được cho như trong hình 2.

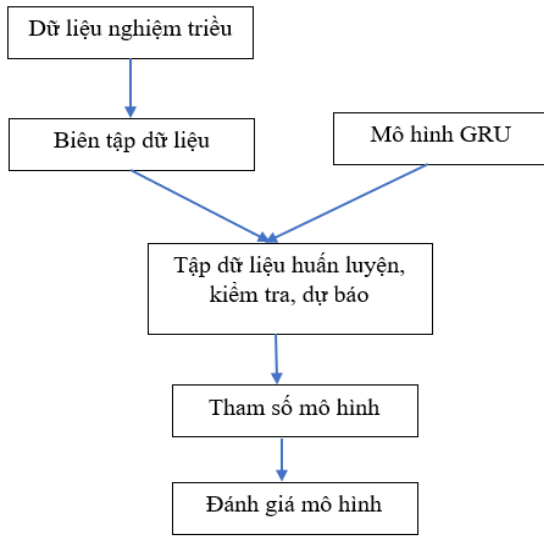
Mạng nơ-ron hồi quy (RNN) là một loại mạng nơ-ron nhân tạo có khả năng xử lý dữ liệu tuần tự, phổ biến trong lĩnh vực xử lý ngôn ngữ tự nhiên và dữ liệu chuỗi thời gian. RNN có khả năng lưu trữ thông tin trong các chuỗi dữ liệu thông qua các kết nối nơ-ron tái phát. Mỗi nút trong RNN nhận đầu vào từ nút trước đó và trạng thái ẩn của nút đó, cho phép mạng học và dự đoán dựa trên ngữ cảnh trước đó. Tuy nhiên, RNN thường gặp vấn đề đối với các chuỗi dài do hiện tượng triệt tiêu gradient.

Mô hình GRU là một biến thể của mạng nơ-ron hồi quy (RNN) được giới thiệu để giải quyết vấn đề đối mặt với hiện tượng triệt tiêu gradient trong việc huấn luyện mạng nơ-ron. GRU sử dụng các cổng (gate) để điều chỉnh luồng thông tin trong quá trình lan truyền ngược, giúp mô hình nhớ thông tin quan trọng và loại bỏ thông tin không quan trọng. Mỗi đơn vị GRU bao gồm cổng cập nhật (update gate) và cổng đặc quyền (reset gate), điều này cho phép mô hình tự động học cách lựa chọn thông tin cần giữ lại và quên đi. Điều này làm giảm được vấn đề vanishing gradient và cải thiện khả năng mô hình học các phụ thuộc dài hạn trong dữ liệu chuỗi. GRU đã chứng minh được hiệu suất trong nhiều ứng dụng như xử lý ngôn ngữ tự nhiên, dịch máy, và dự đoán chuỗi thời gian [20]. Với ưu điểm như trên, nghiên cứu này sử

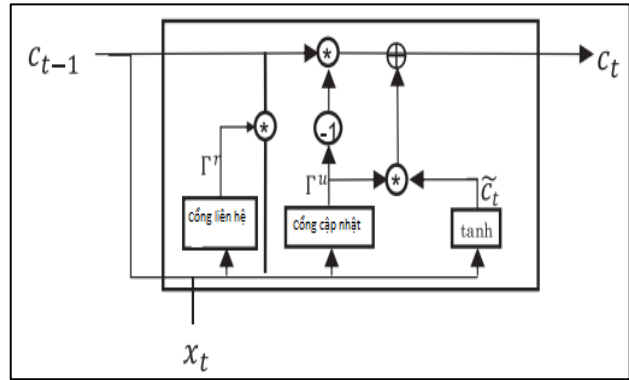


Hình 1. Vị trí Trạm nghiệm triều Vũng Tàu.

dụng mô hình GRU để dự đoán giá trị triều theo chuỗi thời gian. Hiệu quả của mô hình được đánh giá thông qua các đặc trưng thống kê như RMSE, MAE, F-Score, Kappa, NSE và R2 [19, 21]. Mô hình được thực thi theo sơ đồ được biểu diễn như ở hình 3.



Hình 2. Phương pháp nghiên cứu.



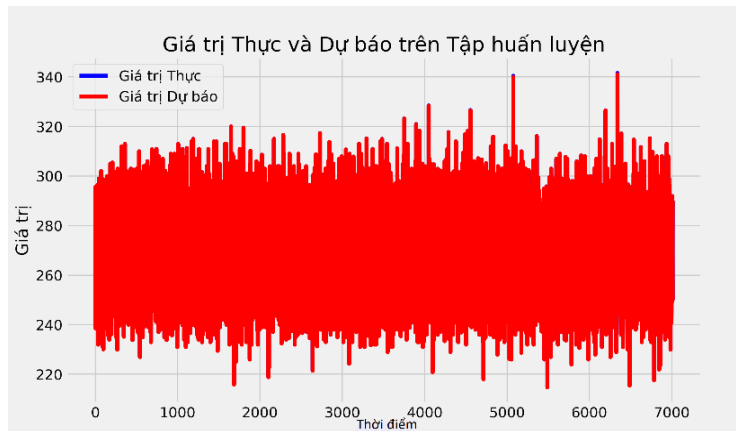
Hình 3. Nguyên lý hoạt động của mô hình GRU.

Trong hình 3, C là tế bào bộ nhớ (hoặc cell state [19]) và x là đầu vào hiện tại ở trạng thái ẩn.

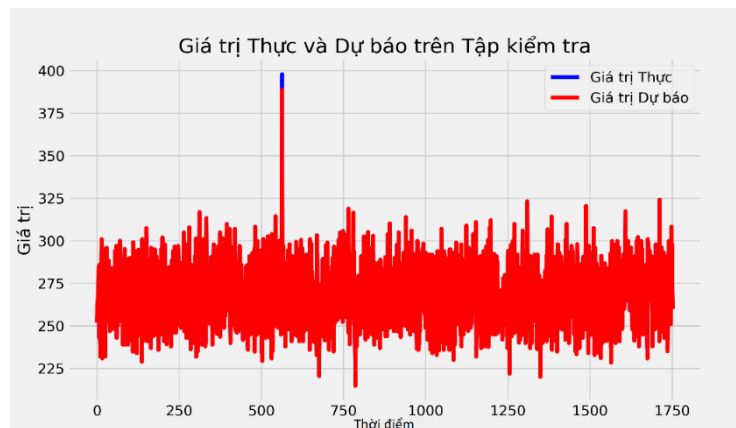
3. Kết quả và thảo luận

Ứng dụng quy trình nghiên cứu như đã mô tả trong phần trên, nghiên cứu này dự báo giá trị triều trong 2 trường hợp trên cơ sở chương trình đã được xây dựng sử dụng ngôn ngữ Python [19]. Các hàm thư viện được sử dụng bao gồm pandas [20], numpy [21], matplotlib [22],... Trường hợp thứ nhất, chia dữ liệu thành tập dữ liệu huấn luyện và tập dữ liệu kiểm tra với tỷ lệ 80% - 20%. Trường hợp thứ 2, chia bộ dữ liệu thành tập dữ liệu huấn luyện, dữ liệu kiểm tra và dữ liệu dự báo với tỷ lệ 60% - 30% - 10%. Trong quá trình huấn luyện mô hình, lựa chọn các thông số bao gồm số epoch = 200; batch-size = 16; trình tối ưu mô hình sử dụng thuật toán Adam. Trong trường hợp thứ nhất, kết quả dự báo với tập dữ liệu huấn luyện được cho như trong hình 4.

Từ hình 4 có thể thấy, mặc dù dữ liệu đầu vào biến thiên khá lớn (có thời điểm từ 310 lên 340 - xấp xỉ cỡ 10% giá trị) nhưng kết quả dự



Hình 4. Kết quả dự báo với tập dữ liệu huấn luyện.



Hình 5. Kết quả dự báo với tập dữ liệu kiểm tra trong trường hợp 1.

báo vẫn khá tốt thể hiện qua sự trùng khớp của đường biểu diễn giá trị thực với đường biểu diễn giá trị dự báo. Tương tự, giá trị dự báo với tập dữ liệu kiểm tra và hàm mất mát trong trường hợp này được cho lần lượt trong hình 5 và hình 6.

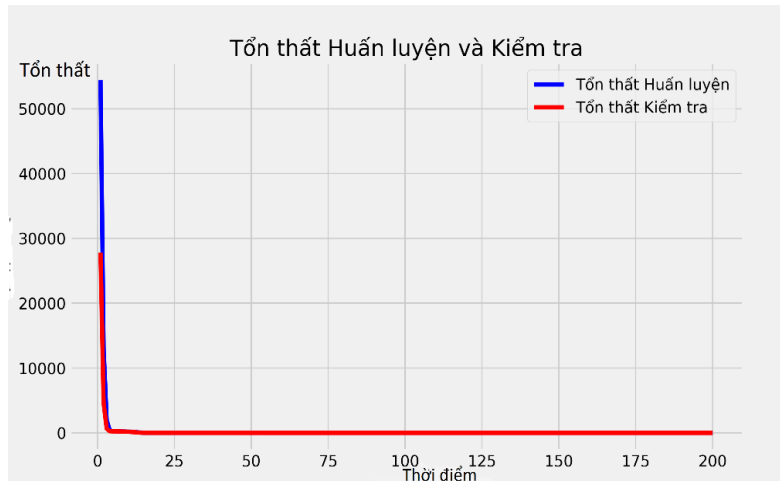
Hình 5 cho thấy, mặc dù có trị đo biến thiên lớn (hoặc là trị đo ngoại lai, thể hiện qua giá trị của trị đo đó đột biến hơn so với các trị đo lân cận) nhưng kết quả dự báo vẫn bám sát so với giá trị thực

chứng tỏ hiệu quả của mô hình GRU trong trường hợp thực nghiệm cụ thể này. Cũng từ hình 5 có thể thấy, trong giá trị dữ liệu của tập dữ liệu đầu vào có giá trị ngoại lai chưa được loại bỏ. Để có được giá trị độ cao mực nước cũng như giá trị tăng lên của mực nước biển một cách chính xác cần thiết phải thực hiện quá trình phát hiện và loại bỏ dữ liệu ngoại lai.

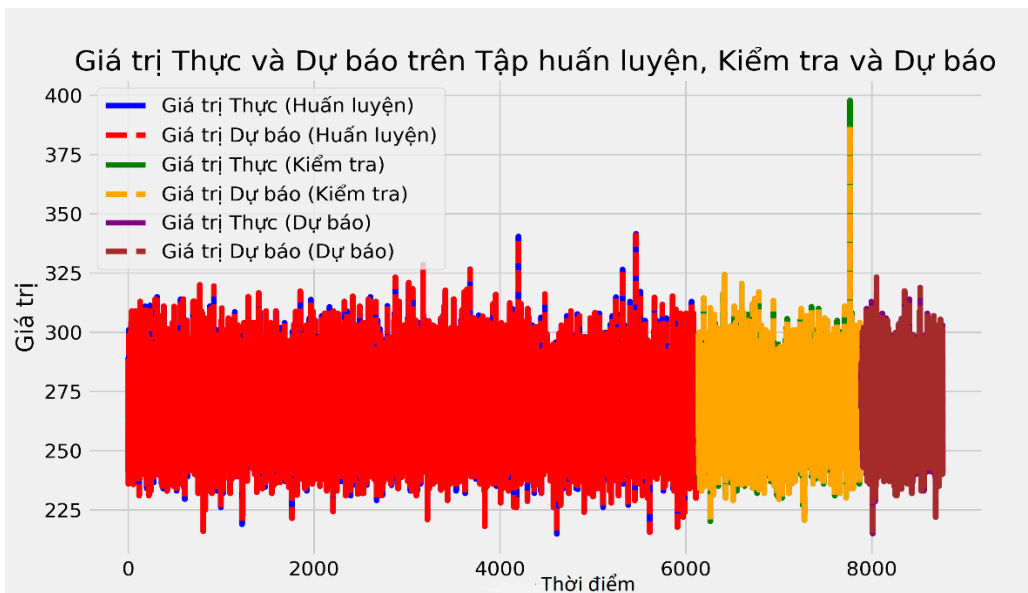
Hình 6 cho thấy, tổn thất khi ước lượng với mô hình GRU gần như bằng không chỉ sau một vài thời điểm (Epoch). Trong quá trình tính này, tại mỗi epoch, chương trình tính lặp 439 lần. Quá trình kiểm tra với tập huấn luyện và tập dữ liệu kiểm tra được tính lặp tương ứng là 220 và 55 lần. Giá trị mất mát trên tập dữ liệu huấn luyện là 0,00072 và trên tập kiểm tra là 0,0496.

Bên cạnh việc thống kê giá trị dự báo, đã tiến hành xác định các tham số thống kê đặc trưng cho mô hình với giá trị RMSE = 2,2 mm; MAE = 0,5 mm; Kappa = 0,98 và F-Score = 0,96. Với kết quả trên cho thấy, hiệu suất dự báo với mô hình GRU rất cao.

Để khẳng định liệu mô hình GRU có thực sự hiệu quả khi ứng dụng trong dự báo dữ liệu nghiệm triều hay không, tiến hành tính thực nghiệm bổ sung với trường hợp thứ 2. Trong trường hợp này, tập dữ liệu dự báo được tạo ra và sử dụng để kiểm chứng mức độ hiệu quả của mô hình. Giá trị thực tế và giá trị dự báo đối với tập dữ liệu, tập dữ liệu kiểm tra và tập dữ liệu dự báo tính toán được với trường hợp 2 cho trong hình 7.

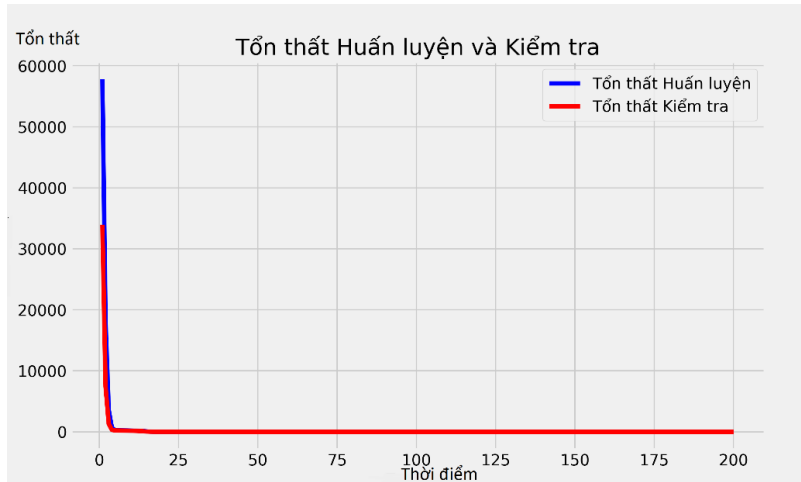


Hình 6. Đường cong mất mát đối với tập dữ liệu huấn luyện và tập dữ liệu kiểm tra.



Hình 7. Đồ thị biểu diễn giá trị thực, giá trị dự báo đối với tập dữ liệu huấn luyện, tập dữ liệu kiểm tra và tập dữ liệu dự báo.

Giá trị hàm mất mát trong trường hợp này được biểu diễn trong hình 8. Quá trình huấn luyện mô hình sử dụng ít lần lặp hơn với 384 lần lặp; Quá trình kiểm tra đối với tập dữ liệu huấn luyện, kiểm tra và dự báo lần lượt là 192, 55 và 28. Giá trị mất mát đối với tập huấn luyện là $9,2318e-04$ và tập dữ liệu kiểm tra tương ứng là 0,0697.



Hình 8. Hàm tổn thất đối với tập dữ liệu huấn luyện và tập dữ liệu kiểm tra trong trường hợp 1.

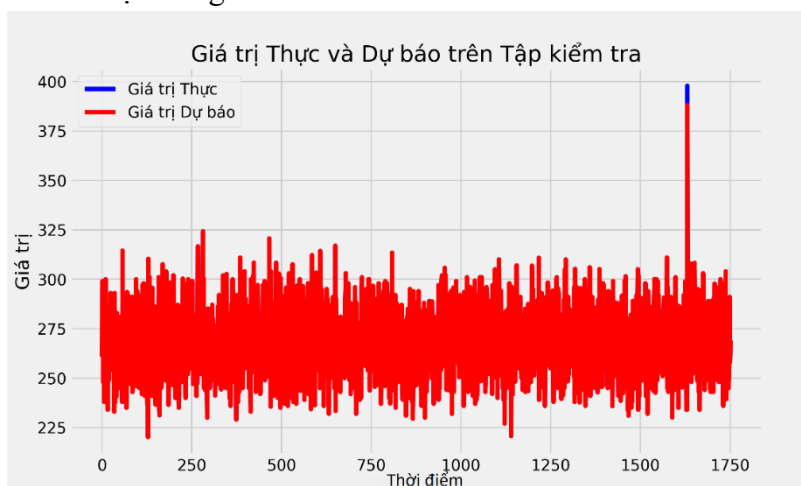
Từ hình 7 có thể thấy, mặc dù bộ dữ liệu huấn luyện đã được thu nhỏ lại (thu nhỏ đi 20% so với trường hợp 1) nhưng hiệu quả dự báo vẫn không hề thay đổi. Thống kê hiệu suất của mô hình trong trường hợp này được thống kê trong bảng 1.

Bảng 1. Thống kê hiệu suất của mô hình.

| | RMSE (mm) | MAE (mm) | Kappa | F-Score |
|------------------------|-----------|----------|-------|---------|
| Tập dữ liệu huấn luyện | 0,17 | 0,1 | 0,99 | 0,948 |
| Tập dữ liệu kiểm tra | 2,6 | 0,11 | 0,991 | 0,968 |
| Tập dữ liệu dự báo | 0,06 | 0,05 | 0,993 | 0,995 |

Lý do giá trị RMSE đối với tập dữ liệu kiểm tra còn lớn do còn tồn tại giá trị ngoại lai trong tập dữ liệu này như được minh họa trong hình 9.

Từ hình 9 một lần nữa cho thấy, cần phải loại bỏ dữ liệu ngoại lai trong tập hợp dữ liệu nghiệm triều đầu vào. Bên cạnh việc thống kê các giá trị như trong bảng 1, nghiên cứu này cũng xác định thêm các tham số NSE, R^2 theo như nghiên cứu [19]. Kết quả đã thu được các giá trị NSE và R^2 đều tiệm cận đến 1 chứng tỏ hiệu suất rất cao của mô hình GRU đã lựa chọn.



Hình 9. Giá trị thực và giá trị dự báo đối với tập dữ liệu kiểm tra trong trường hợp 2.

4. Kết luận

Trong nghiên cứu này, chúng tôi đã tập trung vào phân tích và dự đoán giá trị của mực nước biển sử dụng các mô hình trí tuệ nhân tạo. Kết quả nghiên cứu cho thấy mô hình GRU cho hiệu suất dự báo rất cao thông qua các đặc trưng thống kê của mô hình bao gồm RMSE = 2,2 mm; MAE = 0,5 mm; Kappa = 0,98 và F-Score = 0,96.

Hiệu quả dự báo của mô hình đã được chứng minh thông qua các đặc trưng thống kê của tập dữ liệu dự báo (chiếm 10% tổng số dữ liệu) đều rất cao: RMSE = 0,06 mm; MAE = 0,05 mm; Kappa = 0,993 và F-Score = 0,995. Qua các dữ liệu thống kê trên cho thấy, nhóm nghiên

cứ đã ứng dụng thành công mô hình GRU trong phân tích, dự báo dữ liệu nghiệm triều tại Việt Nam.

Hạn chế của nghiên cứu này đó là chưa tích hợp phát hiện và loại trừ dữ liệu ngoại lai để nâng cao độ chính xác của quá trình dự báo; cũng như chưa xác định được giá trị dâng lên theo chu kỳ (năm) và xu hướng tăng của mực nước biển.

Tuy nhiên, với độ chính xác dự báo như đã đạt được ở trên, hoàn toàn có thể xem xét ứng dụng mô hình GRU trong phân tích dữ liệu nghiệm triều phục vụ các nội dung công việc khác nhau tại Việt Nam với việc sử dụng nhiều bộ dữ liệu nghiệm triều khác để tính toán thực nghiệm.

Đóng góp của tác giả: Xây dựng ý tưởng nghiên cứu: N.G.T., B.N.Q.; Lựa chọn phương pháp nghiên cứu: N.G.T., B.N.Q.; Xử lý số liệu: N.G.T.; Viết bản thảo bài báo: N.G.T., B.N.Q.; Chỉnh sửa bài báo: N.G.T.

Lời cảm ơn: Tập thể tác giả xin gửi lời cảm ơn tới đề tài cấp Bộ Tài nguyên và Môi trường mã số TNMT.2024.02.07 đã cung cấp kinh phí; Đài Khí tượng thủy văn Nam Bộ thành phố Hồ Chí Minh đã cung cấp dữ liệu cho nghiên cứu này.

Lời cam đoan: Tập thể tác giả cam đoan bài báo này là công trình nghiên cứu của tập thể tác giả, chưa được công bố ở đâu, không được sao chép từ những nghiên cứu trước đây; không có sự tranh chấp lợi ích trong nhóm tác giả.

Tài liệu tham khảo

1. Frederikse, T.; Landerer, F.; Caron, L.; Adhikari, S.; Parkes, D.; Humphrey, V.W.; Dangendorf, S.; Hogarth, P.; Zanna, L.; Cheng, L.; Wu, Y.H. The causes of sea-level rise since 1900. *Nature* **2020**, 584, 393–397.
2. Darwin, R.F.; Tol, R.S.J. Estimates of the economic effects of sea level rise. *Environ. Resour. Econ.* **2001**, 19, 113–129.
3. Hamlington, B.D.; Leben, R.R.; Kim, K.Y. Improving sea level reconstructions using non-sea level measurements. *J. Geophys. Res.: Oceans* **2012**, 117(10), 1–14.
4. Antunes, C.; Taborda, R. Sea level at Cascais tide gauge: data, analysis and results. *J. Coastal Res.* **2009**, pp. 218–222.
5. Cipollini, P.; Calafat, F.M.; Jevrejeva, S.; et al. Monitoring sea level in the coastal zone with satellite altimetry and tide gauges. *Surv. Geophys.* **2017**, 38, 33–57.
6. Douglas, B.C. Sea level change in the era of the recording tide gauge. *Int. Geophys.* **2001**, 37–64.
7. Abubakar, A.G.; et al. A Review of modelling approaches on tidal analysis and prediction. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* **2019**, XLII-4/W16, 23–34.
8. French, J.; et al. Combining machine learning with computational hydrodynamics for prediction of tidal surge inundation at estuarine ports. *Procedia IUTAM* **2017**, 25, 28–35.
9. Imani, M.; et al. Daily sea level prediction at Chiayi coast, Taiwan using extreme learning machine and relevance vector machine. *Global Planet. Change* **2018**, 161, 211–221.
10. Bruneau, N.; Polton, J.; Williams, J.; Holt, J. Estimation of global coastal sea level extremes using neural networks. *Environ. Res. Lett.* **2020**, 15(7), 074030.
11. Yang, C.H.; Wu, C.H.; Hsieh, C.M. Long short-term memory recurrent neural network for tidal level forecasting. *IEEE Access* **2020**, 8, 159389–159401.
12. Tur, R.; et al. Sea level prediction using machine learning. *Water* **2021**, 13(24), 3566.
13. Yang, Y.; et al. Multiscale analysis and prediction of sea level in the northern south china sea based on tide gauge and satellite data. *J. Mar. Sci. Eng.* **2023**, 11(6), 1203.

14. Wahle, K.; et al. Detecting anomalous sea-level states in North Sea tide gauge data using an autoassociative neural network. *Nat. Hazards Earth Syst. Sci.* **2023**, 23(2), 415–428.
15. Tín, N.V.; Thịnh, N.N. Nghiên cứu xu thế biến đổi các yếu tố khí hậu, mực nước tại Bà Rịa-Vũng Tàu. *Tạp chí Khí tượng Thủy văn* **2012**, 622, 13–16.
16. Trần, V.C.; Long, B.H. Ảnh hưởng của trường nhiệt độ và biến đổi bất thường của mực nước trong Biển Đông liên quan đến biến đổi khí hậu. *Tạp chí Khoa học và Công nghệ biển* **2016**, 16(3), 255–266.
17. Thủy, N.B.; Tiên, T.Q. Bước đầu nghiên cứu mối liên hệ giữa mực nước biển dâng dị thường tại Tuy Hòa-Phú Yên với hình thế thời tiết. *Tạp chí Khí tượng Thủy văn* **2017**, 687, 15–22.
18. Hai, N.M.; Vinh, V.D. How climate change affected on water level in Ha Long coastal area in the period 1974–2020: results from the mann-kendall test and sen's slope estimate. *VN J. Mar. Sci. Technol.* **2022**, 22(3), 257–269.
19. Thái, T.H.; Khiêm, M.V.; Thủy, N.B.; Hà, B.M.; Ngọc, P.K. Xây dựng mô hình mạng nơ-ron hồi quy dự báo độ cao sóng có nghĩa tại trạm Cồn Cỏ, Quảng Trị, Việt Nam. *Tạp chí Khí tượng Thủy văn* **2022**, EME4, 73–84.
20. Torres, J.F.; Hadjout D.; Sebaa A.; Martínez-Álvarez F.; Troncoso A. Deep learning for time series forecasting: A survey. *Big Data* **2020**, 9(1), 3–21.
21. Phong, D.V.; Trọng, N.G.; Chiên, N.V.; Thành, N.H.; Hà, L.L.; Quân, N.V.; Quang, P.N. Phân tích chuyên dịch thẳng đứng vô Trái đất sử dụng hàm ANN từ kết quả xử lý chuỗi dữ liệu GNSS theo thời gian. *Tạp chí Khí tượng Thủy văn* **2023**, 752, 41–50.
22. Trục tuyến: <https://www.python.org/>.
23. Trục tuyến: <https://anaconda.org/anaconda/pandas>.
24. Trục tuyến: <https://anaconda.org/anaconda/numpy>.
25. Trục tuyến: <https://anaconda.org/conda-forge/matplotlib>.

Analyzing tidal data sequences using Gated Recurrent Unit (GRU)

Nguyen Gia Trong^{1,2*}, Bui Ngoc Quy³

¹ Hanoi University of Mining and Geology, nguyengiatrong@humg.edu.vn

² Geodesy and Environment research group, Hanoi University of Mining and Geology

³ VNU - Central Institute for Natural Resources and Environmental Studies (VNU-CRES); quybnres@vnu.edu.vn

Abstract: This study applied the GRU (Gated Recurrent Unit) model, a type of Recurrent Neural Network (RNN), to analyze tidal data collected at the Vung Tau tide gauge station from January 1, 1999, to December 31, 2022. Experimental results show that the statistical performance of the model is very promising: RMSE = 2.2 mm; MAE = 0.5 mm; Kappa = 0.98, and F-Score = 0.96. Additionally, to verify the effectiveness of the model, the dataset was divided into training, testing, and prediction datasets. Statistical results indicate that the GRU model can reliably predict tidal values using statistical features of the prediction dataset, such as RMSE = 0.06mm, MAE = 0.05 mm. The study also highlights the necessity of implementing outlier detection and elimination for the input dataset to achieve high prediction accuracy.

Keywords: RNN; GRU; Tidal gauges data; Data time series; AI.