

FORECASTING STREAMFLOW USING A HYBRID CNN-BILSTM MODEL WITH ATTENTION MECHANISM: A CASE STUDY FOR TUYEN QUANG HYDROPOWER RESERVOIR

Nguyen Thi Hoai Thu*, Trinh Trong Nam

School of Electrical and Electronic Engineering - Hanoi University of Science and Technology

ARTICLE INFO		ABSTRACT
Received:	15/4/2025	Forecasting the streamflow to hydropower reservoirs plays a crucial role in optimizing the operation and management of hydropower plants, ensuring efficient energy production and rational water resource utilization. This study proposes a hybrid deep learning model that combines a convolutional neural network with a bidirectional long short-term memory network integrated with an attention mechanism for hydropower inflow forecasting. To evaluate the forecasting capability in real-time scenarios, the data was resampled at a daily frequency, and predictions were made for the next 1 day, 1 week, and 1 month. The proposed model was compared with various deep learning models, including CNN, LSTM, CNN-LSTM, and CNN-BiLSTM. The results demonstrate that the proposed model outperforms the others, particularly in the 1-week forecast, achieving MAE, RMSE, and N-RMSE values of 176.52 m ³ /s, 427.29 m ³ /s, and 7.47%, respectively. These findings highlight the model's potential to support more informed and efficient reservoir operation decisions. By doing so, it contributes to improving the reliability of water resource management and optimizing hydropower output under dynamic operating conditions.
Revised:	14/6/2025	
Published:	15/6/2025	
KEYWORDS		
Forecasting		
Resampling		
Streamflow		
Hybrid model		
CNN-BiLSTM-Attention		

DỰ BÁO LƯU LƯỢNG NƯỚC ĐẾN HỒ CHỨA NHÀ MÁY THỦY ĐIỆN TUYẾN QUANG BẰNG MÔ HÌNH LẠI CNN-BiLSTM KẾT HỢP CƠ CHẾ CHÚ Ý

Nguyễn Thị Hoài Thu*, Trịnh Trọng Nam

Trường Điện - Điện tử, Đại học Bách khoa Hà Nội

THÔNG TIN BÀI BÁO		TÓM TẮT
Ngày nhận bài:	15/4/2025	Dự báo lưu lượng nước về hồ chứa thủy điện đóng vai trò quan trọng trong việc tối ưu hóa vận hành và quản lý các nhà máy thủy điện, đảm bảo sản xuất năng lượng hiệu quả và sử dụng tài nguyên nước hợp lý. Nghiên cứu này đề xuất một mô hình học sâu lai kết hợp mạng nơ-ron tích chập với mạng nơ-ron có bộ nhớ dài ngắn hai chiều kết hợp cơ chế chú ý để dự báo lưu lượng nước thủy điện. Để đánh giá khả năng dự báo trong các kịch bản thời gian thực, dữ liệu được tái lấy mẫu theo chu kỳ ngày, và thực hiện dự báo cho 1 ngày, 1 tuần và 1 tháng tới. Mô hình đề xuất được so sánh với các mô hình học sâu bao gồm CNN, LSTM, CNN-LSTM, CNN-BiLSTM. Kết quả cho thấy mô hình đề xuất vượt trội, đặc biệt trong dự báo tuần tới với chỉ số MAE, RMSE và N-RMSE lần lượt là 176,52 m ³ /s, 427,29 m ³ /s và 7,47%. Nhờ đó, mô hình có thể hỗ trợ hiệu quả cho việc ra quyết định vận hành hồ chứa, góp phần nâng cao độ tin cậy trong quản lý tài nguyên nước và tối ưu hóa sản lượng phát điện từ thủy điện trong điều kiện vận hành biến động.
Ngày hoàn thiện:	14/6/2025	
Ngày đăng:	15/6/2025	
TỪ KHÓA		
Dự báo		
Tái lấy mẫu		
Lưu lượng nước đến		
Mô hình lai		
CNN-BiLSTM-Cơ chế chú ý		

DOI: <https://doi.org/10.34238/tnu-jst.12584>

* Corresponding author. Email: thu.nguyenthihoai@hust.edu.vn

1. Đặt vấn đề

Một trong những phương thức sản xuất điện năng phổ biến trên thế giới là thông qua các nhà máy thủy điện. Do chi phí vận hành thấp so với các nguồn năng lượng khác nên ở những quốc gia có tiềm năng về thủy văn, nguồn năng lượng này được sử dụng phổ biến. Việt Nam là một trong những quốc gia có tỷ trọng năng lượng thủy điện lớn trong cơ cấu nguồn năng lượng. Dự báo đến năm 2030, thủy điện có công suất 29.356 MW, chiếm khoảng 19,5% tổng sản lượng điện của cả nước [1]. Do đó, việc dự báo lưu lượng nước về hồ đóng vai trò then chốt trong công tác quản lý tài nguyên nước, hỗ trợ ứng phó với biến đổi khí hậu và nâng cao hiệu quả vận hành các nhà máy thủy điện.

Trong thời gian qua, nhiều nghiên cứu đã được thực hiện nhằm áp dụng các mô hình dự báo khác nhau, với mục tiêu cải thiện độ chính xác và hiệu quả của công tác dự báo lưu lượng. Một số mô hình thống kê như mô hình tự hồi quy (AR) [2], mô hình hồi quy tự động trung bình trượt (ARMA) [3], mô hình trung bình trượt tích hợp tự hồi quy có tính mùa vụ (SARIMA) [4] được sử dụng rộng rãi để dự báo lưu lượng. Tuy nhiên, các mô hình truyền thống hạn chế trong việc nắm bắt các đặc điểm phi tuyến của dữ liệu. Hiện nay, các mô hình học máy đang được nghiên cứu và sử dụng rộng rãi, do khả năng phân tích dữ liệu lớn với nhiều tham số, đồng thời khắc phục được hạn chế của các mô hình thống kê. Các mô hình học máy được sử dụng phổ biến trong dự báo dòng chảy như mạng nơ-ron nhân tạo (ANN) [5], máy học cực trị (ELM) [6], máy vectơ liên quan (RVM) [7]. Mặc dù các mô hình học máy đã được chứng minh hiệu quả vượt trội trong nhiều mô hình và bài toán dự báo, nhưng chúng vẫn tồn tại một số nhược điểm nhất định như phụ thuộc rất nhiều vào chất lượng và khối lượng dữ liệu đầu vào. Nếu dữ liệu bị thiếu, nhiễu hay mất cân bằng thì mô hình có thể học sai hoặc không tổng quát hóa tốt. Chính vì vậy, nhiều nghiên cứu gần đây đã ứng dụng các mô hình mạng nơ-ron học sâu như mạng nơ-ron tích chập (CNN) [4], mạng nơ-ron hồi quy với bộ nhớ dài ngắn (LSTM) [8]. Kết quả cho thấy các mô hình học sâu này mang lại hiệu quả trong việc dự báo lưu lượng nước. Mặc dù vậy, trong một số trường hợp, các mô hình dự báo đơn lẻ không thể đạt được hiệu quả cao. Do đó, việc kết hợp hai hay nhiều mô hình trở thành một giải pháp tiềm năng nhằm cải thiện chất lượng dự báo. Việc áp dụng mô hình lai giúp tận dụng những ưu điểm của từng mô hình đơn lẻ. Trong nghiên cứu [6], mô hình lai CNN-LSTM cho kết quả sai số tốt hơn so với các mô hình CNN và LSTM riêng biệt. Tuy nhiên, các nghiên cứu hiện tại vẫn chủ yếu sử dụng các mô hình lai đơn giản mà chưa tích hợp thêm các lớp phức tạp hơn.

Nhà máy thủy điện Tuyên Quang nằm trên dòng sông Gâm, thuộc thị trấn Na Hang, huyện Na Hang, tỉnh Tuyên Quang, Việt Nam. Nhà máy được khởi công xây dựng vào năm 2002 và khánh thành sau 5 năm, với 3 tổ máy và công suất thiết kế là 342 MW. Hồ thủy điện Tuyên Quang có tổng diện tích trên 8000 ha có dung tích chứa từ 1,5 đến 2 tỷ m³. Đây là nhà máy có vai trò quan trọng trong hệ thống điện Việt Nam, ngoài nhiệm vụ đảm bảo phát điện thì hồ chứa còn làm nhiệm vụ cung cấp nước cho sản xuất nông nghiệp và phòng chống lũ. Vì vậy, việc xây dựng mô hình dự báo lưu lượng nước về hồ thủy điện Tuyên Quang với độ chính xác cao là rất cần thiết.

Mục tiêu của nghiên cứu này là xây dựng một mô hình dự báo lưu lượng nước về hồ thủy điện Tuyên Quang trong các khoảng thời gian khác nhau: ngắn hạn (1 ngày), trung hạn (7 ngày), dài hạn (30 ngày) nhằm hỗ trợ công tác vận hành và điều tiết hồ chứa hiệu quả. Dựa trên các phân tích ở trên, bài báo này đề xuất xây dựng mô hình lai kết hợp mạng nơ-ron tích chập (CNN) và mạng hồi tiếp hai chiều (BiLSTM) kết hợp với cơ chế chú ý. Dữ liệu được thu thập từ nhà máy thủy điện Tuyên Quang, sau khi xử lý, sẽ được chia thành tập huấn luyện và kiểm tra. Các tập dữ liệu này sẽ được đưa vào mô hình CNN-BiLSTM-Attention để huấn luyện và đánh giá. Hiệu quả của mô hình được kiểm chứng trên tập kiểm tra thông qua các chỉ số đánh giá sai số phổ biến, đồng thời so sánh với các mô hình đơn và lai khác để chứng minh tính ưu việt của mô hình đề xuất.

2. Phương pháp luận

2.1. Tiền xử lý dữ liệu

Dữ liệu đầu vào của mô hình dự báo lưu lượng nước về hồ thủy điện là lưu lượng nước trong quá khứ và thu được từ các trạm quan trắc. Trong thực tế, dữ liệu này thường có kích thước lớn,

không đồng nhất, vì vậy cần được xử lý trước khi đưa vào các mô hình dự báo, học sâu. Trong quá trình tiền xử lý, dữ liệu có thể được làm sạch bằng cách xử lý các giá trị bị thiếu, loại bỏ hoặc điều chỉnh giá trị ngoại lai. Ngoài ra, dữ liệu cần phải được chuẩn hóa để đảm bảo tính đồng nhất nhằm tối ưu hiệu suất của mô hình.

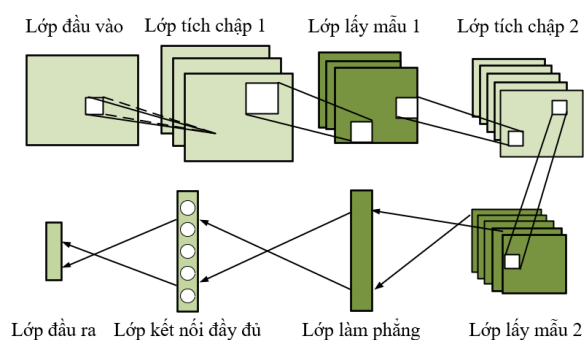
- Xử lý dữ liệu khuyết thiếu: trong nghiên cứu này, dữ liệu bị thiếu (NaN) có thể được thay thế bằng giá trị trung bình của hai giá trị lân cận.

- Chuẩn hóa dữ liệu: để đảm bảo dữ liệu đầu vào có cùng đơn vị đo lường và tránh ảnh hưởng của các giá trị lớn hoặc nhỏ bất thường, phương pháp chuẩn hóa Min-Max được sử dụng để đưa dữ liệu về khoảng $[0, 1]$. Việc chuẩn hóa được thực hiện theo biểu thức (1):

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (1)$$

Trong đó, X là giá trị gốc của dữ liệu, $X_{\min}$ và $X_{\max}$ lần lượt là giá trị nhỏ nhất và lớn nhất của tập dữ liệu, X' là giá trị sau khi chuẩn hóa, nằm trong khoảng $[0, 1]$.

2.2. Mô hình CNN



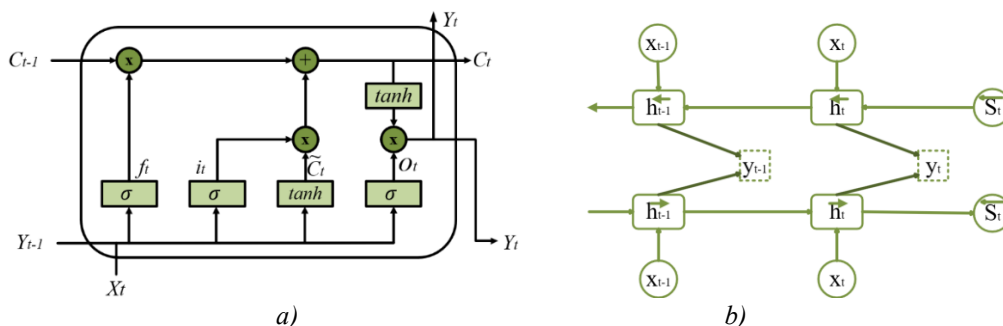
Hình 1. Cấu trúc của mô hình CNN

Mạng nơ-ron tích chập (CNN) là một mô hình học sâu phổ biến, thường được sử dụng trong các tác vụ như nhận diện hình ảnh, thị giác máy tính. Bên cạnh đó, CNN ngày càng được ứng dụng rộng rãi trong phân tích và xử lý dữ liệu dạng chuỗi thời gian nhờ khả năng học đặc trưng vượt trội. Mô hình CNN được cấu thành từ hai thành phần chính: lớp tích chập và lớp lấy mẫu, được minh họa trong Hình 1. Mỗi lớp tích chập sử dụng nhiều kernel tích chập, và quá trình tính toán được diễn tả thông qua phương trình (2).

$$l_t = \tanh(x_t \times k_t + b_t) \quad (2)$$

Trong đó, l_t là đầu ra sau khi thực hiện phép tích chập; x_t là vector đầu vào tại thời điểm t ; k_t biểu thị trọng số của bộ lọc; b_t là giá trị độ lệch; còn $\tanh$ là hàm kích hoạt phi tuyến của hàm tiếp tuyến hyperbolic.

2.3. Mô hình BiLSTM



Hình 2. Cấu trúc của một tế bào LSTM đơn lẻ (a) và cấu trúc của BiLSTM

Trong lĩnh vực dự báo, mô hình BiLstm được sử dụng rộng rãi nhờ khả năng nắm bắt các mối quan hệ dài hạn. Mô hình Bi-LSTM bao gồm nhiều đơn vị LSTM hoạt động theo cả hai hướng để tích hợp thông tin từ quá khứ và tương lai. Cấu trúc của BiLSTM được minh họa trong Hình 2. Mỗi tế bào đơn lẻ bao gồm nhiều công khác nhau, và hoạt động của các công này được đặc tả thông qua các phương trình toán học trong [9].

2.4. Cơ chế Attention

Cơ chế Attention giúp xác định mức độ quan trọng của các bước thời gian trong chuỗi đầu ra của lớp BiLSTM. Mỗi bước thời gian được gán một trọng số thích nghi dựa trên mức độ tương thích với một vector ngữ cảnh được học. Trọng số được tính toán dựa trên sự kết hợp giữa các đặc trưng ẩn của chuỗi dữ liệu tại mỗi thời điểm, với mục tiêu tập trung vào những thông tin có giá trị cao và bỏ qua các thông tin ít liên quan hơn.

Tại mỗi thời điểm t , điểm số chú ý cho mỗi đặc trưng đầu vào tính toán theo công thức (3):

$$S_t = \tanh(X_t \cdot W + b) \quad (3)$$

Trong đó, X_t là chuỗi dữ liệu vào tại thời điểm t ; W là ma trận trọng số; b là hệ số điều chỉnh.

Trọng số chú ý mỗi đặc trưng đầu vào được tính toán thông qua softmax theo công thức (4):

$$\alpha_k = \text{softmax}(S_t \cdot u) \quad (4)$$

Trong đó, u là vector trọng số học; α_k là vector trọng số chú ý của đặc trưng đầu vào x_k .

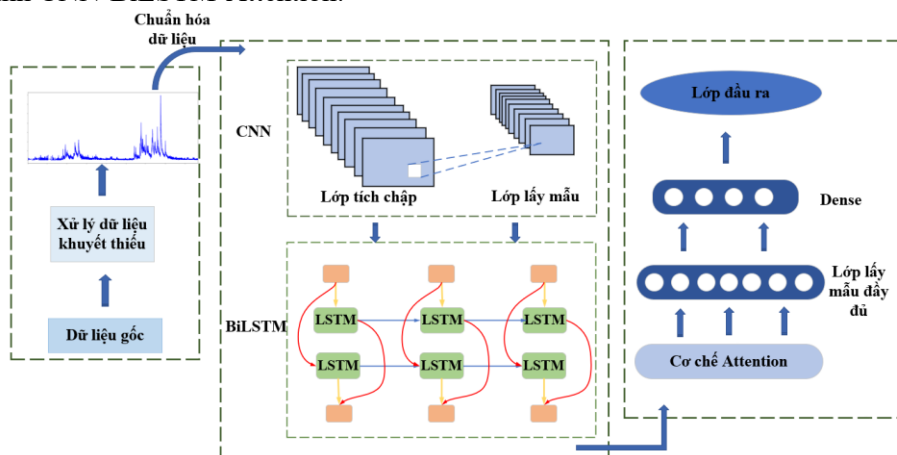
Đầu ra của cơ chế attention tại thời điểm t được tính bằng cách tổng hợp các đặc trưng đầu vào với trọng số chú ý tương ứng, được mô tả thông qua phương trình (5):

$$z_t = \sum_{k=1} \alpha_k \cdot x_k \quad (5)$$

Trong đó, z_t là đầu ra của cơ chế attention tại thời điểm t .

2.5. Mô hình đề xuất

Bài báo đề xuất một mô hình lai CNN-BiLSTM kết hợp với cơ chế Attention, giúp mô hình tập trung vào các phân quan trọng của dữ liệu và cải thiện hiệu suất dự báo. Hình 3 thể hiện cấu trúc của mô hình CNN-BiLSTM-Attention.



Hình 3. Sơ đồ cấu trúc mô hình CNN-BiLSTM-Attention

Quá trình xây dựng mô hình dự báo bao gồm các bước:

Bước 1. Thu thập và tiền xử lý dữ liệu: dữ liệu được thu thập từ nhà máy thủy điện Tuyên Quang, sau đó được xử lý các giá trị thiếu, ngoại lai và được chuẩn hóa bằng phương pháp Min-Max để dữ liệu đầu vào phù hợp với yêu cầu của mô hình học sâu. Sau đó, dữ liệu được chia thành các tập huấn luyện và kiểm tra với tỷ lệ 70:30.

Bước 2. Xây dựng mô hình CNN-BiLSTM-Attention: các lớp CNN sẽ được sử dụng để trích xuất đặc trưng không gian, trong khi BiLSTM giúp mô hình học các quan hệ phụ thuộc theo thời gian, và cơ chế Attention giúp mô hình tập trung vào các phần quan trọng của dữ liệu đầu vào.

Bước 3. Huấn luyện mô hình: mô hình được huấn luyện với hàm mất mát Sai số bình phương trung bình (MSE) và bộ tối ưu Adam. Các tham số như số lớp, số lượng nơ-ron được điều chỉnh qua các thử nghiệm thủ công, thử nghiệm nhiều lần để tìm ra bộ tham số tối ưu cho mô hình.

Bước 4. Đánh giá mô hình: sau khi huấn luyện, mô hình được đánh giá trên tập kiểm tra thông qua các chỉ số như MAE, RMSE, N-RMSE để đo lường độ chính xác và khả năng dự báo của mô hình. Các kết quả này sẽ được so sánh với các mô hình học sâu khác để xác định hiệu quả và tính ưu việt của mô hình CNN-BiLSTM-Attention.

Trong nghiên cứu này, ngôn ngữ Python và thư viện TensorFlow/Keras được sử dụng để thực hiện các bước tiền xử lý dữ liệu và xây dựng mô hình. Các chương trình được triển khai và các phép toán được thực hiện trên Google Colab, sử dụng GPU T4 với RAM 12GB, 2560 lõi CUDA.

2.6. Các chỉ số đánh giá mô hình

Bài báo này đánh giá hiệu quả của các mô hình dự báo thông qua ba chỉ số sai số phổ biến, bao gồm sai số tuyệt đối trung bình (MAE), sai số căn phương trung bình (RMSE), sai số RMSE chuẩn hóa (N-RMSE)

$$MAE = \frac{1}{N} \sum_{i=1}^n |Y_i - y_i| \quad (6)$$

$$RMSE = \sqrt{\frac{1}{N} \times \sum (Y_i - y_i)^2} \quad (7)$$

$$N-RMSE = \frac{RMSE}{\max(Y_i) - \min(y_i)} \times 100\% \quad (8)$$

trong đó, N là tổng số mẫu trong tập dữ liệu; Y_i là giá trị thực tế tại thời điểm thứ i ; y_i là giá trị dự báo tương ứng tại thời điểm thứ i .

Để đánh giá độ chính xác của mô hình, chỉ số N-RMSE được sử dụng, với các mức phân loại như sau: mô hình được coi là xuất sắc khi N-RMSE dưới 10%, tốt khi nằm trong khoảng từ 10% đến dưới 20%, trung bình nếu N-RMSE từ 20% đến dưới 30%, và kém khi đạt 30% trở lên.

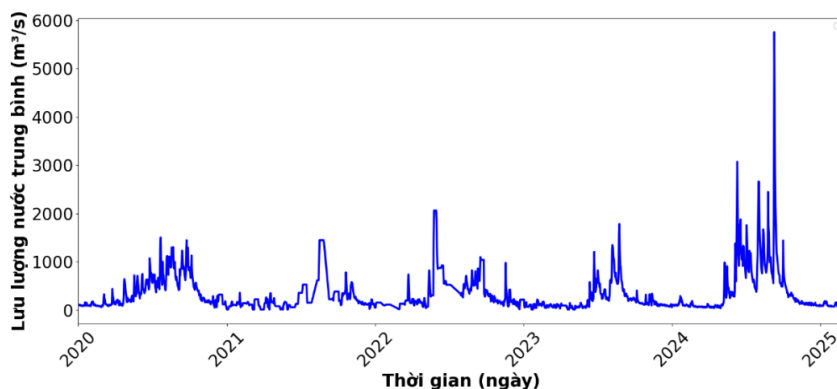
3. Kết quả và thảo luận

3.1. Thu thập và xử lý dữ liệu

Dữ liệu sử dụng trong nghiên cứu là lưu lượng nước về hồ chứa thủy điện Tuyên Quang, được thu thập từ Trang thông tin vận hành hồ chứa của Tập đoàn Điện lực Việt Nam (EVN) trong khoảng thời gian từ ngày 1/1/2020 đến ngày 1/3/2025, với tần suất lấy mẫu là 1 giờ [10]. Tập dữ liệu có tổng cộng 43091 điểm dữ liệu (Bảng 1). Có thể thấy được, tập dữ liệu đang có giá trị khuyết thiếu, và được thay thế bằng giá trị trung bình đã được trình bày trong mục 2.1.1. Dữ liệu đã được tái lấy mẫu thành chu kỳ 24 giờ bằng cách tính giá trị trung bình theo ngày. Lưu lượng nước về trung bình ngày trong giai đoạn từ năm 2020 đến 2025 được thể hiện trong Hình 4.

Bảng 1. Thống kê của tập dữ liệu trước và sau xử lý

Lưu lượng nước về	Trước xử lý	Sau xử lý
Count	43091	45288
Mean	339,29	338,78
Std	421,65	416,49
Min.	4,3	4,3
Max.	6966	6966



Hình 4. Lưu lượng nước về trung bình ngày của hồ chứa thủy điện Tuyên Quang trong giai đoạn từ năm 2020 đến năm 2025

3.2. Tinh chỉnh siêu tham số các mô hình dự báo

Các siêu tham số của mô hình đề xuất và các mô hình so sánh được sử dụng trong bài báo này được trình bày trong Bảng 2. Các tham số huấn luyện của mô hình đã được chọn với các giá trị khác nhau và lấy giá trị tốt nhất.

Bảng 2. Các tham số của mô hình đề xuất và mô hình so sánh

Siêu tham số của mô hình	Models					Mô hình đề xuất
	CNN	LSTM	BiLSTM	CNN-LSTM	CNN-BiLSTM	
Conv1D_1	96	-	-	72	54	72
MaxPooling_1	2	-	-	2	2	2
Conv1D_2	48	-	-	-	-	-
MaxPooling_2	2	-	-	-	-	--
BiLSTM_1	-	-	128	-	72	64
BiLSTM_2	-	-	64	-	36	32
LSTM_1	-	112	-	128	-	-
LSTM_2	-	54	-	64	-	-
Attention	-	-	-	-	-	Có
Dropout	0,3	0,3	0,1	0,2	0,2	0,2
Dense_1	32	64	64	32	32	32
Dense_2	16	32	32	16	16	4
Activation	ReLU	ReLU	ReLU	ReLU	ReLU	ReLU
Optimizer	Adam	Adam	Adam	Adam	Adam	Adam

3.3. Đánh giá và so sánh mô hình đề xuất với các mô hình khác

Trong nghiên cứu này, dữ liệu được chia theo tỷ lệ 70:30, với kích thước dữ liệu đầu vào là 24. Mô hình CNN-BiLSTM-Attention sẽ được so sánh với các mô hình dự báo phổ biến, bao gồm CNN, LSTM, BiLSTM, CNN-LSTM, CNN-BiLSTM. Bảng 3 trình bày các sai số dự báo cùng với thời gian chạy của các mô hình. Ba sai số được dùng để đánh giá kết quả dự báo là MAE, RMSE và N-RMSE, thực hiện dự báo cho 1 ngày, 1 tuần và 1 tháng tới.

Từ kết quả dự báo đơn bước trong Bảng 3 có thể thấy rằng mô hình đề xuất đạt hiệu quả tốt nhất với giá trị RMSE và N-RMSE thấp nhất có giá trị lần lượt là 251,90 m³/s, 4,40%. Mặc dù chỉ số MAE của mô hình đề xuất cao hơn một chút so với mô hình CNN-BiLSTM, nhưng sự chênh lệch này là không đáng kể.

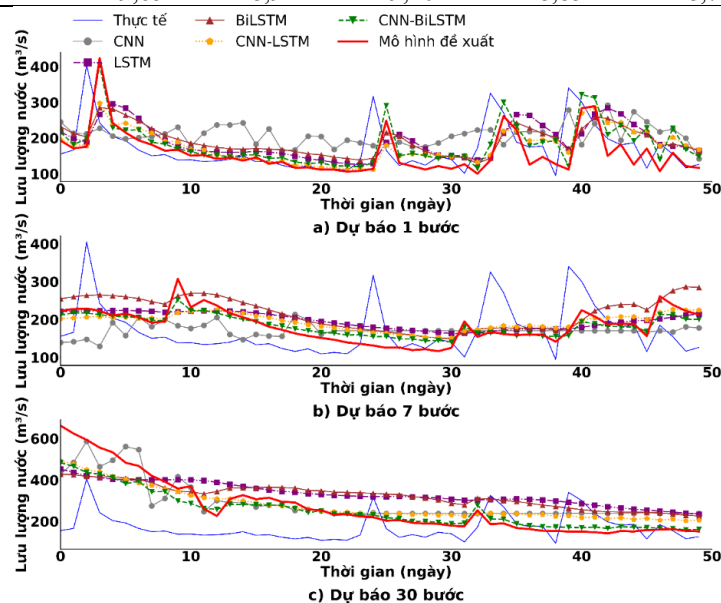
Đối với dự báo đa bước, sai số của các mô hình đều có xu hướng tăng, do khoảng thời gian dự báo dài hơn làm gia tăng độ bất định và khó khăn trong việc nắm bắt xu hướng dữ liệu. Tuy nhiên, mô hình đề xuất tiếp tục duy trì hiệu quả cao với các chỉ số MAE, RMSE và N-RMSE thấp nhất

trong cả dự báo trung hạn (theo tuần) và dài hạn (theo tháng). Kết quả này cho thấy tính ổn định cũng như khả năng tổng quát hóa vượt trội của mô hình so với các mô hình còn lại.

Hình 5 biểu diễn kết quả dự báo của các mô hình với dự báo cho ngày, tuần, tháng tới. Ở Hình 5a, có thể thấy các mô hình có khả năng bám sát với giá trị thực tế rất tốt, đặc biệt là mô hình đề xuất gần như trùng khớp với đường thực tế. Tuy nhiên, mô hình CNN thể hiện sai lệch rõ rệt nhất trong số các mô hình, đặc biệt tại các điểm đỉnh và đáy, cho thấy khả năng phản ứng kém với các biến động ngắn hạn. Điều này cho thấy CNN gặp khó khăn trong việc nắm bắt các đặc trưng tạm thời hoặc thay đổi đột ngột trong chuỗi thời gian. Tiếp theo ở Hình 5b và 5c, sự sai lệch giữa các mô hình với dữ liệu thực tế ngày càng tăng theo số bước dự báo. Các mô hình đều thể hiện độ trễ, đặc biệt tại những đoạn biến động. Trong đó, mô hình đề xuất tiếp tục thể hiện tốt nhất xu hướng chính của dữ liệu thực tế trong các mô hình.

Bảng 3. Kết quả sai số và thời gian chạy của các mô hình, tương ứng với dự báo 1, 7 và 30 bước

Số bước	Chỉ số	CNN	LSTM	BiLSTM	CNN-LSTM	CNN-BiLSTM	Mô hình đề xuất
1	MAE	183,70	108,93	107,88	89,80	87,92	88,06
	RMSE	427,37	303,20	303,92	284,10	264,32	251,90
	N-RMSE	7,47	5,30	5,31	4,97	4,62	4,40
7	MAE	243,12	184,43	187,44	199,18	178,87	176,52
	RMSE	533,37	445,69	439,57	483,38	465,28	427,29
	N-RMSE	9,33	7,79	7,69	8,45	8,13	7,47
30	MAE	306,67	272,22	268,36	265,08	251,88	225,11
	RMSE	552,18	510,01	520,29	506,14	497,97	469,08
	N-RMSE	9,65	8,92	9,10	8,85	8,71	8,20



Hình 5. Kết quả dự báo của các mô hình với bước lấy mẫu 1 ngày:
(a) Dự báo 1 bước, (b) Dự báo 7 bước, (c) Dự báo 30 bước

4. Kết luận

Nghiên cứu này đã phát triển một mô hình lai kết hợp giữa CNN và BiLSTM tích hợp cơ chế Attention nhằm dự báo lưu lượng nước về hồ chứa thủy điện Tuyên Quang. Dữ liệu được tải lấy mẫu theo chu kỳ 24 giờ để phục vụ cho việc phân tích xu hướng dài hạn, với các kịch bản dự báo cho 1 ngày, 1 tuần, 1 tháng tới. Các mô hình như CNN, LSTM, BiLSTM, CNN-LSTM, và CNN-BiLSTM được sử dụng để so sánh với mô hình đề xuất. Kết quả cho thấy mô hình kết hợp Attention đạt độ chính xác cao vượt trội trong cả dự báo ngắn hạn, trung hạn và dài hạn. Trong điều kiện

thực tế tại Việt Nam, để áp dụng mô hình này trong dự báo lưu lượng nước về hồ chứa, cần đảm bảo một số điều kiện. Về điều kiện cần, dữ liệu đầu vào phải liên tục, chính xác và đủ dài, với bước lấy mẫu ổn định theo giờ hoặc ngày. Ngoài ra, cần có sẵn dữ liệu lịch sử chất lượng cao để huấn luyện mô hình. Còn điều kiện đủ, cần có hạ tầng tính toán đủ mạnh, phần mềm phù hợp. Trong các nghiên cứu tiếp theo, có thể mở rộng việc thử nghiệm mô hình trên nhiều lưu vực khác hoặc trong điều kiện thời tiết, địa hình phức tạp hơn sẽ giúp đánh giá tính khái quát và khả năng ứng dụng rộng rãi của phương pháp.

TÀI LIỆU THAM KHẢO/ REFERENCES

- [1] Prime Minister of Vietnam, "National Power Development Plan VIII," (in Vietnamese), May 18, 2023. [Online]. Available: <https://xaydungchinh sach.chinhphu.vn/toan-van-quy-hoach-phat-trien-dien-luc-quoc-gia-11923051616315244.htm>. [Accessed Apr. 15, 2025].
- [2] Ö. Terzi and G. Ergin, "Forecasting of monthly river flow with autoregressive modeling and data-driven techniques," *Neural Comput. & Applic.*, vol. 25, no. 1, pp. 179–188, Jul. 2014, doi: 10.1007/s00521-013-1469-9.
- [3] K. Mohammadi, H. R. Eslami, and R. Kahawita, "Parameter estimation of an ARMA model for river flow forecasting using goal programming," *Journal of Hydrology*, vol. 331, no. 1–2, pp. 293–299, Nov. 2006, doi: 10.1016/j.jhydrol.2006.05.017.
- [4] R. Maiti, B. G. Menon, and A. Abraham, "Ensemble empirical mode decomposition based deep learning models for forecasting river flow time series," *Expert Systems with Applications*, vol. 255, Dec. 2024, Art. no. 124550, doi: 10.1016/j.eswa.2024.124550.
- [5] F. Fathian, S. Mehdizadeh, A. K. Sales, and M. J. S. Safari, "Hybrid models to improve the monthly river flow prediction: Integrating artificial intelligence and non-linear time series models," *Journal of Hydrology*, vol. 575, pp. 1200–1213, Aug. 2019, doi: 10.1016/j.jhydrol.2019.06.025.
- [6] J. Wang, X. Li, R. Wu, X. Mu, B. Baiyinbaoligao, J. Wei, J. Gao, D. Yin, X. Tao, and K. Xu, "A runoff prediction approach based on machine learning, ensemble forecasting and error correction: A case study of source area of Yellow River," *Journal of Hydrology*, vol. 658, Sep. 2025, Art. no. 133190, doi: 10.1016/j.jhydrol.2025.133190.
- [7] J. Flake, T. K. Moon, M. McKee, and J. H. Gunther, "Application of the relevance vector machine to canal flow prediction in the Sevier River Basin," *Agricultural Water Management*, vol. 97, no. 2, pp. 208–214, Feb. 2010, doi: 10.1016/j.agwat.2009.09.010.
- [8] T. T. Tran, H. N. Pham, Q. B. Pham, N. P. Nguyen, D. P. Nguyen, N. D. Liem, and P. T. Khuat, "Application of Long Short-Term Memory neural network for time series prediction of flow rate at My Thuan hydrology station, Tien river," *Science and Technology Development Journal – Natural Sciences*, vol. 6, no. 1, pp. 1884–1896, Feb. 2022, doi: 10.32508/stdjns.v6i1.1129.
- [9] T. H. T. Nguyen and Q. B. Phan, "Hourly day ahead wind speed forecasting based on a hybrid model of EEMD, CNN-Bi-LSTM embedded with GA optimization," *Energy Reports*, vol. 8, pp. 53–60, Nov. 2022, doi: 10.1016/j.egy.2022.05.110.
- [10] Vietnam Electricity (EVN), "Hydropower Reservoir Water Levels," (in Vietnamese), 2025. [Online]. Available: <https://www.evn.com.vn/c3/thong-tin-ho-thuy-dien/Muc-nuoc-cac-ho-thuy-dien-117-123.aspx>. [Accessed Apr. 25, 2025].