

DỰ BÁO MỨC NƯỚC SÔNG CẨM, THÀNH PHỐ HẢI PHÒNG BẰNG MÔ HÌNH MẠNG NƠ-RON LSTM

Hồ Việt Hùng

Trường Đại học Thủy lợi

Tóm tắt: Sông Cẩm là sông lớn thuộc địa phận Hải Phòng, giữ vị trí trọng yếu về kinh tế, quốc phòng và văn hóa không chỉ của Hải Phòng mà của miền Bắc nước ta. Gần đây nhiều khu đô thị lớn, hiện đại được xây dựng bên bờ sông Cẩm. Vì vậy, dự báo chính xác mực nước sông Cẩm sẽ góp phần quan trọng trong việc phòng chống ngập lụt, đảm bảo an toàn đời sống nhân dân và phát triển kinh tế, xã hội. Theo đó, tác giả bài báo này đã thiết lập một mô hình Long Short-Term Memory Neural Network (LSTM), một dạng đặc biệt của Mạng nơ-ron hồi quy (Recurrent Neural Network - RNN) để dự báo mực nước sông Cẩm tại trạm Cửa Cẩm, Hải Phòng. Mô hình dự báo chỉ cần dữ liệu đầu vào là mực nước thực đo tại các trạm thủy văn và hải văn trong khu vực nghiên cứu. Lượng mưa tại các trạm: Cao Kênh, Kiến An, Phù Liễn, Cửa Cẩm có hệ số tương quan thấp nên các chuỗi số liệu này không được sử dụng cho mô hình. Hệ số Nash (Nash Sutcliffe Efficiency), Sai số bình phương trung bình (Root Mean Squared Error), Sai số tuyệt đối trung bình (Mean Absolute Error) đã được sử dụng để đánh giá sai số trị số dự báo. Kết quả dự báo có độ chính xác cao, chất lượng dự báo là đủ độ tin cậy. Do đó, có thể áp dụng mô hình này để dự báo mực nước sông Cẩm và các sông khác ở Hải Phòng.

Từ khóa: Mạng nơ-ron hồi quy (RNN), LSTM, dự báo mực nước, sông Cẩm, Hải Phòng.

Summary: The Cam River is a big river in Hai Phong, holding an important position related to economy, national defense and culture not only of Hai Phong but also of Northern Vietnam. Recently, many large and modern urban centers have been built on the banks of the Cam River. Therefore, accurately forecasting the water levels in the Cam River will make an important contribution to flood prevention, ensuring the safety of people's lives and socio-economic development. Accordingly, the author of this article has set up a Long Short-Term Memory Neural Network (LSTM) model, a special type of the Recurrent Neural Network (RNN), to predict the water levels of the Cam River at Cua Cam station in Hai Phong. The input data of the forecast model is only the water levels measured at the hydrological stations in the study area. Rainfall at stations: Cao Kanh, Kien An, Phu Lien, Cua Cam have low correlation coefficients, so these data series are not used for the model. Nash Sutcliffe Efficiency, Root Mean Squared Error, Mean Absolute Error were used to evaluate the errors of the forecast values. The forecast results are highly accurate, predictive quality is sufficiently reliable. Therefore, this model can be applied to forecast the water levels of the Cam River and other rivers in Hai Phong.

Keywords: Recurrent Neural Network (RNN), LSTM, water level forecast, Cam River, Hai Phong.

1. GIỚI THIỆU CHUNG

Sông Cẩm là sông lớn, quan trọng vào loại bậc nhất Hải Phòng, giữ vị trí trọng yếu về kinh tế, quốc phòng và văn hóa không chỉ của địa

phương mà của miền Bắc nước ta. Toàn bộ sông Cẩm thuộc địa phận Hải Phòng, bắt đầu từ ngã ba Hợp Thành (từ vị trí nhập lưu của sông Kinh Thầy - Kinh Môn) đến nhập lưu vào sông Bạch

Ngày nhận bài: 21/01/2021

Ngày thông qua phản biện: 04/02/2021

Ngày duyệt đăng: 08/02/2021

Đằng để đổ ra biển qua cửa Nam Triệu. Sông Cấm nối với sông Lạch Tray qua sông Tam Bạc. Sông có chiều rộng tương đối lớn, chỗ hẹp nhất khoảng 200 m, chỗ rộng nhất 700 m (từ cầu Kiền ra phía biển). Cao độ đáy sông ở chỗ sâu nhất là -12 m.

Gần đây nhiều khu đô thị lớn, hiện đại được xây dựng bên bờ sông Cấm. Vì vậy, dự báo chính xác mực nước sông Cấm sẽ góp phần quan trọng trong việc phòng chống ngập lụt, đảm bảo an toàn đời sống nhân dân và phát triển kinh tế, xã hội.

Hiện nay ở Việt Nam cũng như trên thế giới, các mô hình mạng nơ-ron nhân tạo (Artificial Neural Network - ANN) ngày càng được ứng dụng nhiều trong các lĩnh vực khác nhau, trong đó có ngành Thủy lợi. Mô hình ANN được sử dụng để dự báo mực nước, lưu lượng dòng chảy trong sông nhằm cảnh báo lũ lụt, hạn hán [1], [3], [4], hoặc dự báo mực nước ở các cống tưới tiêu kết hợp trong hệ thống thủy lợi [5]. Đây là các dự báo thời hạn cực ngắn hay thời hạn ngắn, phục vụ trực tiếp cho công tác điều hành hệ thống hoặc phòng chống thiên tai. Mô hình ANN khác với các mô hình thủy văn, thủy lực truyền thống là không yêu cầu một lượng lớn các dữ liệu đầu vào như địa hình, địa mạo, các hệ số và kích thước của các công trình trên sông, kênh... Trong trường hợp không có đủ tài liệu địa hình, địa mạo, biển – đại dương và tùy theo yêu cầu thực tế, các mô hình ANN có thể được áp dụng để dự báo các yếu tố thủy văn, thủy lực của dòng chảy. Các mô hình này dựa trên mối quan hệ giữa các dữ liệu đầu vào và đầu ra để học các quy luật của dòng chảy từ đó thực hiện các dự báo.

Trên thế giới, các mô hình ANN đã được sử dụng để dự báo mực nước sông từ năm 1990. Sau đó, các thuật toán của mô hình ANN đã được cải tiến nhằm tăng độ chính xác trong dự báo [1]. Asaad Y. Shamseldin (2010) [7] đã sử dụng mô hình ANN để dự báo lưu lượng dòng chảy sông Nile xanh ở Sudan. Trong nghiên cứu

của mình, tác giả đã sử dụng giá trị lưu lượng dòng chảy trung bình ngày của sông Nile xanh đo tại Eldeim gần biên giới Sudan - Ethiopia và dữ liệu lượng mưa trung bình ngày trong khu vực cho giai đoạn bốn năm 1992-1995 để dự báo. Chen, J.F. và nnk (2014) [8] đã dự báo lưu lượng dòng chảy đến hồ Hòa Bình, Việt Nam bằng mô hình có sử dụng thuật toán Cuckoo Search. Sung và nnk (2017) [10] đã sử dụng mô hình mạng nơ-ron hồi quy (Recurrent Neural Network - RNN) để dự báo mực nước trên sông Anyangcheon, Hàn Quốc, trước 1 giờ và 3 giờ, với dữ liệu đầu vào là mực nước sông trong khu vực nghiên cứu. Veintimilla-Reyes (2016) [11] đã dự báo lưu lượng dòng chảy sông Tomebamba, Ecuador bằng mô hình ANN với thời gian dự báo lên đến 3 ngày. Dữ liệu đầu vào của mô hình là lượng mưa và lưu lượng đo tại các trạm khí tượng, thủy văn trên lưu vực sông.

Ngoài ra, còn có các nghiên cứu ở trong nước như: Lê Xuân Hiền, Hồ Việt Hùng (2018) đã ứng dụng mô hình mạng nơ-ron Long Short-Term Memory (LSTM) để dự báo mực nước trên các sông ở Hải Phòng. Các tác giả đã sử dụng dữ liệu đầu vào là mực nước theo giờ để dự báo mực nước sông từ 1 giờ đến 5 giờ trong tương lai [1]. Lê Xuân Hiền và Hồ Việt Hùng (2018) đã xây dựng một mô hình toán dựa trên mạng nơ-ron hồi quy để dự báo lưu lượng lũ sông Đà tại Lai Châu trước 1 ngày [2] và dự báo lưu lượng dòng chảy sông Hồng tại trạm Sơn Tây trước 1 ngày, 2 ngày và 3 ngày [3]. Mô hình mà các tác giả đề xuất không đòi hỏi các dữ liệu về địa hình, địa mạo, chỉ cần các dữ liệu về lưu lượng dòng chảy. Lê Xuân Hiền và nnk đã sử dụng mô hình LSTM để dự báo lưu lượng nước về hồ Hòa Bình trước 6 giờ, 12 giờ và 18 giờ [9], [6]. Hồ Việt Hùng (2019) [4] đã thiết lập một mô hình mạng nơ-ron hồi quy để dự báo lưu lượng dòng chảy sông Hồng và đỉnh lũ tại Hà Nội trước 1 ngày, 2 ngày và 3 ngày. Hồ Việt Tuấn và Hồ Việt Hùng (2019) [5] đã ứng dụng mô hình LSTM để dự báo mực nước tại hạ lưu

công - âu thuyền Cầu Cát trên sông Thái Bình, thời gian dự báo là 6h, 12h, 18h và 24h. Những nghiên cứu trên đây cho thấy rằng, các mô hình ANN và RNN ngày càng được ứng dụng rộng rãi trong lĩnh vực thủy văn, thủy lực để dự báo thời hạn ngắn các trị số mực nước hoặc lưu lượng dòng chảy trong sông.

Theo đó, tác giả bài báo này sẽ trình bày sau đây việc thiết lập một mô hình toán dựa trên mạng nơ-ron hồi quy (RNN) để dự báo mực nước sông Cẩm tại trạm Cửa Cẩm, Hải Phòng. Mô hình dự báo chỉ sử dụng dữ liệu đầu vào là mực nước thực đo tại các trạm thủy văn, hải văn ở Hải Phòng. Đây là dự báo thời hạn ngắn: 1h, 3h, 6h.

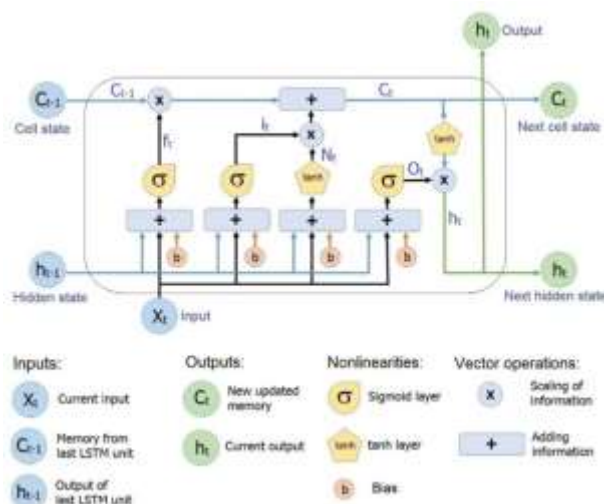
2. PHƯƠNG PHÁP NGHIÊN CỨU VÀ DỮ LIỆU

Trong nghiên cứu này, thư viện phần mềm mã nguồn mở Keras, các thư viện Numpy, Pandas, cùng với ngôn ngữ lập trình Python 3.6 đã được sử dụng để thiết lập một mô hình LSTM.

2.1. Mạng nơ-ron LSTM và các dữ liệu đầu vào mô hình

Mạng nơ-ron RNN có chứa các vòng lặp bên trong, cho phép thông tin được lưu lại. Mạng nơ-ron LSTM cũng như RNN có thể giải quyết các bài toán có sự phụ thuộc dài hạn (long-term dependency). Các mô hình RNN dựa trên mối quan hệ giữa các dữ liệu đã có để học các quy luật trong quá khứ, từ đó đưa ra các dự báo cho tương lai. Các mô hình LSTM xuất hiện lần đầu tiên vào năm 1997, cho đến nay đã được cải tiến nhiều và phổ biến rộng rãi [5], [6].

Mạng nơ-ron LSTM có cấu trúc dạng chuỗi, gồm nhiều mô đun được lặp lại, nó có 4 tầng tương tác với nhau một cách đặc biệt như Hình 1 dưới đây [5].



Hình 1: Cấu trúc của một mô đun LSTM [5]

Mỗi mô đun LSTM gồm có: trạng thái tế bào (cell state); các cổng (gate) [5]. Cell state chạy xuyên suốt từ mô đun này sang mô đun khác, giúp thông tin được truyền đi dễ dàng; các cổng là nơi sàng lọc thông tin. Trong mỗi mô đun có 3 cổng và 4 tầng. Đầu tiên là tầng cổng quên f_t (forget gate layer), nó quyết định thông tin nào cần loại bỏ từ Cell state. Đầu vào của tầng này là h_{t-1} (giá trị đầu ra tại thời điểm $t-1$) và x_t (dữ liệu đầu vào hiện tại); đầu ra của tầng này là f_t , một số trong khoảng từ 0 đến 1 cho mỗi số trong Cell state C_{t-1} .

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

Trong đó: σ là hàm sigmoid, W_f và b_f lần lượt là trọng số và tham số của tầng cổng quên.

Hai tầng tiếp theo sẽ quyết định thông tin lưu vào Cell state và cập nhật giá trị cho nó, đó là tầng cổng vào i_t (input gate layer) và một tầng tanh N_t (tanh layer).

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$N_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

$$C_t = f_t * C_{t-1} + i_t * N_t \quad (4)$$

Trong đó: C_{t-1} và C_t là Cell state lần lượt ở thời điểm $t-1$ và t ; W_i và b_i lần lượt là trọng số và tham số của tầng cổng vào. W_c và b_c là trọng

số và tham số của Cell state; $\tanh$ là hàm hyperbolic tangent.

Cuối cùng là tầng công ra o_t (output gate layer), giá trị đầu ra (h_t) sẽ được quyết định bởi Cell state muốn xuất ra.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t * \tanh(C_t) \quad (6)$$

Trong đó: W_o và b_o lần lượt là trọng số và tham số của tầng công ra.

Để phục vụ cho mô hình LSTM, các dữ liệu đã được thu thập gồm có: lượng mưa và mực nước thực đo theo giờ ở các trạm khí tượng, thủy văn trên các sông: Thái Bình, Kinh Thầy, Cấm, Lạch Tray và đảo Hòn Dấu. Tác giả bài báo này đã tham khảo các tài liệu, báo cáo của các nghiên cứu trước đây về dòng chảy trên các sông ở Hải Phòng trong mùa lũ, đánh giá số liệu của trận lũ lớn tháng 8/1996 và tháng 8/2002. Tác giả đã so sánh mực nước lũ thiết kế tại trạm Cửa Cấm (sông Cấm) trong trường hợp có nước dâng, theo số liệu của năm 2011, với mực nước lũ thiết kế theo công văn số

527/BNN-TCTL ngày 6/3/2012 của Bộ Nông nghiệp và PTNT về việc thỏa thuận qui hoạch phòng lũ chi tiết các tuyến sông có đề trên địa bàn thành phố Hải Phòng đến năm 2020, thấy rằng các số liệu mực nước mùa lũ năm 2011 là phù hợp cho mô hình dự báo.

Trên cơ sở dữ liệu hiện có, các số liệu được sử dụng cho bài báo này bao gồm: lượng mưa ở Hải Phòng và mực nước theo giờ được đo ở các trạm Cửa Cấm trên sông Cấm, Cao Kênh trên sông Kinh Thầy, Kiến An trên sông Lạch Tray và đảo Hòn Dấu, trong thời gian từ 0 giờ ngày 14/7/2011 đến 23 giờ ngày 30/9/2011 (1896 số liệu). Các số liệu thực đo đã được xử lý và đánh giá bằng hệ số tương quan r khi đối chiếu lượng mưa và mực nước tại các trạm với mực nước trạm Cửa Cấm (Bảng 1). Lượng mưa tại các trạm: Cao Kênh, Kiến An, Phù Liễn, Cửa Cấm có hệ số tương quan thấp, như vậy, chuỗi mực nước ở Cửa Cấm không có tương quan rõ ràng với lượng mưa trong khu vực. Do đó, các chuỗi số liệu lượng mưa không được sử dụng cho mô hình dự báo.

Bảng 1: Hệ số tương quan (r) của các chuỗi dữ liệu

	Cao Kênh (s. Kinh Thầy)	Kiến An (s. Lạch Tray)	Hòn Dấu (Biển Đông)	Cửa Cấm (s. Cấm)
Hệ số r	0,986	0,988	0,917	1,0

Như vậy, các chuỗi dữ liệu mực nước thực đo tại các trạm Cao Kênh, Kiến An, Hòn Dấu và Cửa Cấm được sử dụng làm dữ liệu đầu vào cho mô hình. Tập dữ liệu đầu vào được chia thành 3 phần với các mục đích khác nhau. Phần thứ nhất là chuỗi 1536 số liệu được đo đạc từ 0h ngày 14/7/2011 đến 23h ngày 15/9/2011, dùng để hiệu chỉnh mô hình (training), phần này chiếm 81% dữ liệu. Phần thứ hai là chuỗi số liệu đo từ 0h ngày 16/9/2011 đến 23h ngày 28/9/2011 được dùng để kiểm định mô hình

(testing). Phần thứ ba là chuỗi số liệu từ 0h ngày 29/9/2011 đến 23h ngày 30/9/2011 được sử dụng làm dữ liệu đầu vào để chạy thử mô hình nhằm kiểm tra một lần nữa (validating). Như vậy, 19% dữ liệu được dùng để kiểm định mô hình. Sơ đồ mạng lưới sông và vị trí các trạm thủy văn được thể hiện trong Hình 2.



Hình 2: Mạng lưới sông ở Hải Phòng [1]

2.2. Thiết kế mô hình LSTM

Tác giả đã chạy mô hình với nhiều bộ thông số khác nhau để chọn một bộ thông số cho kết quả dự báo tốt nhất. Số lớp ẩn của mô hình LSTM là một. Các thông số khác của mô hình được tóm tắt dưới đây

- Số lượng unit trong mỗi mô đun: 20; 30
- Hệ số học (Learning rate): 0,001
- Bộ tối ưu hóa (Optimizer): Adam
- Số lần lặp tối đa (Epoch): 5000
- Các kỹ thuật được sử dụng: Early Stopping, Regularizer.
- Mục tiêu dự báo: mực nước tại Cửa Cấm ở thời điểm $(t+1)$, $(t+3)$, $(t+6)$
- Dữ liệu đầu vào: mực nước tại Cao Kênh, Kiên An, Hòn Dấu, Cửa Cấm. Có 3 trường hợp dự báo được thống kê cụ thể trong Bảng 2.

Bảng 2: Các trường hợp dự báo

Biến phụ thuộc	Biến độc lập
Mực nước Cửa Cấm ở thời điểm $(t+1)$	Mực nước Cao Kênh, Kiên An, Hòn Dấu, Cửa Cấm ở thời điểm (t) , $(t-1)$, $(t-2)$... $(t-5)$
Mực nước Cửa Cấm ở thời điểm $(t+3)$	Mực nước Cao Kênh, Kiên An, Hòn Dấu, Cửa Cấm ở thời điểm (t) , $(t-1)$, $(t-2)$... $(t-8)$
Mực nước Cửa Cấm ở thời điểm $(t+6)$	Mực nước Cao Kênh, Kiên An, Hòn Dấu, Cửa Cấm ở thời điểm (t) , $(t-1)$, $(t-2)$... $(t-11)$

2.3. Phương pháp đánh giá sai số

Để đánh giá sai số của trị số dự báo và chất lượng dự báo, tác giả sử dụng Sai số bình phương trung bình (RMSE), Sai số tuyệt đối trung bình (MAE), Sai số trung bình (Bias) theo Thông tư số 42/2017/TT-BTNMT, ngày 23/10/2017, Quy định kỹ thuật đánh giá chất lượng dự báo, cảnh báo thủy văn [12], và hệ số Nash (Nash Sutcliffe Efficiency – NSE) [1], được tính toán theo các công thức dưới đây. Mô hình dự báo có độ chính xác cao khi trị số NSE gần bằng 1.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2} \quad (7)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n (|P_i - O_i|) \quad (8)$$

$$Bias = \frac{1}{n} \sum_{i=1}^n (P_i - O_i) \quad (9)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (10)$$

Trong đó: O_i , $\bar{O}$ và P_i lần lượt là trị số thực đo,

trị số thực đo trung bình và trị số dự báo tương ứng thứ i ; n là số lần phát báo.

Để xác định độ tin cậy của trị số dự báo, sai số cho phép (S_{cf}) và độ lệch chuẩn của yếu tố dự báo (σ_2) được tính toán bằng công thức (11) và (12), theo [12].

$$S_{cf} = 0,674\sigma_2 \quad (11)$$

$$\sigma_2 = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}} \quad (12)$$

Trong đó: S_{cf} là sai số cho phép; σ_2 là độ lệch chuẩn của yếu tố dự báo; Y_i là trị số của yếu tố trong dãy số tính toán; $\bar{Y}$ là chuẩn của dãy số tính toán; n là số số hạng trong dãy số tính toán.

Chất lượng dự báo trị số được xác định là "đủ độ tin cậy" khi sai số trị số dự báo nhỏ hơn hoặc bằng sai số cho phép của yếu tố dự báo.

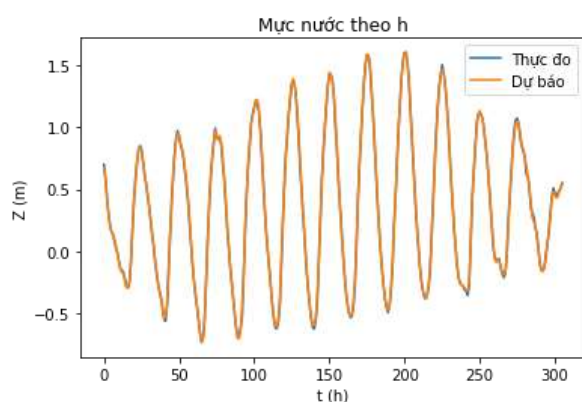
3. KẾT QUẢ NGHIÊN CỨU

3.1. Kết quả kiểm định mô hình

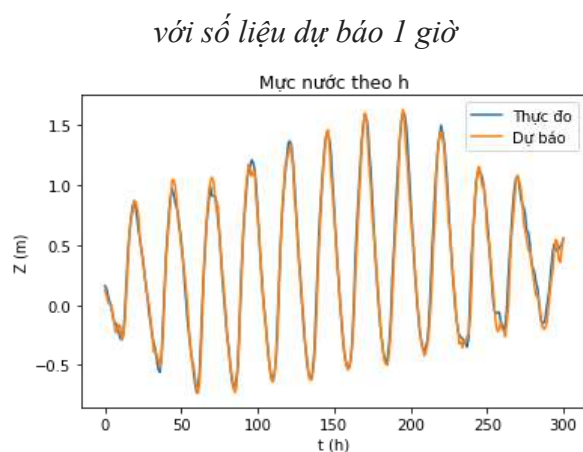
Mô hình LSTM đã được kiểm định bằng phần dữ liệu thứ hai. Kết quả kiểm định tốt nhất của ba trường hợp dự báo được thể hiện trong Bảng 3 và các Hình 3, 4, 5. Bảng 3 cho thấy rằng, Hệ số Nash dao động từ 99,9% (dự báo 1h) đến 89,7% (dự báo 6h); Sai số bình phương trung bình dưới 20 cm cho tất cả các trường hợp dự báo, khi dự báo 1h sai số là không đáng kể, khi dự báo xa hơn, độ chính xác (hệ số NSE) giảm xuống. Có thể thấy mực nước lớn nhất khi dự báo 6h thấp hơn thực đo. Kết quả kiểm định mô hình là tốt. Chất lượng dự báo là "đủ độ tin cậy" vì các sai số đều nhỏ hơn sai số cho phép. Mô hình đủ điều kiện và đạt yêu cầu để tiến hành dự báo.

Bảng 3: Kết quả kiểm định mô hình dự báo mực nước

Số bước thời gian dự báo	Số chuỗi dữ liệu	Số lượng unit	Số lần lặp	RMSE (m)	MAE (m)	Bias (m)	Sai số cho phép (m)	NSE (%)
t+1	4	20	920	0,023	0,017	0,003	0,409	99,9
t+3	4	20	1284	0,078	0,060	-0,028	0,412	98,3
t+6	4	30	730	0,197	0,156	0,138	0,415	89,7

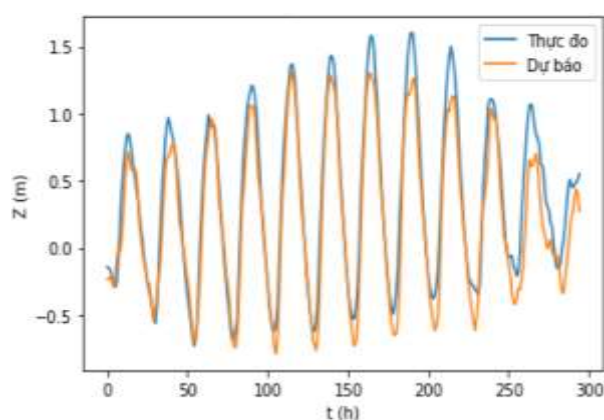


Hình 3: So sánh mực nước thực đo



Hình 4: So sánh mực nước thực đo với số liệu dự báo 3 giờ

Sau quá trình hiệu chỉnh và kiểm định, các thông số của mô hình đã được chọn cho từng trường hợp dự báo. Các mô hình này được ghi lại thành 4 files dưới định dạng *.h5 phục vụ cho từng trường hợp dự báo khác nhau.



Hình 5: So sánh số liệu thực đo với mực nước dự báo 6 giờ

3.2. Kết quả dự báo thử nghiệm

Tác giả tiếp tục kiểm định mô hình bằng cách dự báo thử mực nước tại trạm Cửa Cấm trước 1h, 3h và 6h với dữ liệu đầu vào là 6; 9 hoặc 12 số liệu mực nước tại thời điểm bất kỳ trong các ngày 29/9 và 30/9/2011. Để tiến hành dự báo thử nghiệm cho từng trường hợp, tác giả cho chạy các file định dạng *.h5 trong môi trường Python 3.6 có sử dụng các thư viện keras, numpy, pandas, sklearn. Phần dữ liệu thứ ba đã được sử dụng cho mục đích thử nghiệm này. Đây là tập dữ liệu độc lập, chưa được sử dụng trước đây nhằm đảm bảo tính khách quan của việc dự báo.

Kết quả dự báo được so sánh với số liệu thực đo nhằm đánh giá mô hình thêm một lần nữa (xem Bảng 4). Bảng 4 cho thấy: sai số tuyệt đối dao động từ 1 cm đến 24 cm; sai số lớn nhất xảy ra khi dự báo 6 bước thời gian. Mực nước dự báo có xu thế thấp hơn thực đo.

Bảng 4: Kết quả dự báo mực nước ở thời điểm bất kỳ

Ngày, Tháng, Năm đo MN	Giờ đo MN	Thời gian dự báo	Dữ liệu đầu vào (số)	MN thực đo (m)	MN dự báo (m)	Sai số (m)	Sai số cho phép (m)
29/9/2011	17:00	1 giờ	6	-0,13	-0,12	0,01	0,409
29/9/2011	19:00	3 giờ	9	-0,09	0,03	0,12	0,412
30/9/2011	4:00	6 giờ	12	0,55	0,31	0,24	0,415

Tiếp theo, tác giả tiến hành dự báo mực nước lớn nhất ở cuối tháng 9, xuất hiện lúc 5h ngày 30/9/2011. Chi tiết về kết quả dự báo được thể

hiện trong Bảng 5. Sai số tuyệt đối trong khoảng từ 1 cm đến 20 cm. Kết quả này là chấp nhận được. Các sai số đều nhỏ hơn sai số cho phép,

tức là chất lượng dự báo đủ độ tin cậy. Như vậy, mô hình LSTM đã được kiểm định xong, sau hai lần dự báo thử nghiệm với sai số nhỏ,

Bảng 5: Kết quả dự báo mực nước lớn nhất (lúc 5h ngày 30/9/2011)

Ngày, Tháng, Năm đo MN	Giờ đo MN	Thời gian dự báo	Dữ liệu đầu vào (số)	MN thực đo (m)	MN dự báo (m)	Sai số (m)	Sai số cho phép (m)
30/9/2011	5:00	1 giờ	6	0,59	0,60	0,01	0,409
30/9/2011	5:00	3 giờ	9	0,59	0,48	0,11	0,412
30/9/2011	5:00	6 giờ	12	0,59	0,39	0,20	0,415

4. KẾT LUẬN

Trong bài báo này, tác giả đã trình bày phương pháp nghiên cứu để xây dựng một mô hình LSTM, trình bày kết quả thiết lập và kiểm định mô hình nhằm dự báo mực nước sông Cẩm, Hải Phòng, với dữ liệu đầu vào là mực nước thực đo theo giờ tại 4 trạm: Cao Khê, Cửa Cẩm, Kiến An, Hòn Dấu. Lượng mưa thực đo trong phạm vi Hải Phòng ảnh hưởng không đáng kể đến kết

quả dự báo, do đó không được sử dụng trong bài báo này. Mô hình mà tác giả đề xuất có thể sử dụng để dự báo mực nước tại trạm Cửa Cẩm cho 1, 3 và 6 bước thời gian, với dữ liệu đầu vào là số liệu mực nước của 6, 9 hoặc 12 bước thời gian trước đó, tùy theo mục đích dự báo. Khi mô hình này được ứng dụng, sai số trị số dự báo nhỏ, chất lượng dự báo là đủ độ tin cậy. Do đó, có thể áp dụng mô hình này để dự báo mực nước sông Cẩm và các sông khác ở Hải Phòng.

TÀI LIỆU THAM KHẢO

- [1] Lê Xuân Hiền, Hồ Việt Hùng (2018). “Ứng dụng mạng Long Short-Term Memory (LSTM) để dự báo mực nước tại trạm Quang Phục và Cửa Cẩm, Hải Phòng, Việt Nam”. Tạp chí Thủy lợi và môi trường số 62 (tháng 9/2018), Đại học Thủy lợi. Trang 9-16.
- [2] Hồ Việt Hùng, Lê Xuân Hiền, Giha Lee (2018). “xây dựng mô hình mạng nơ-ron hồi quy dựa trên phần mềm mã nguồn mở để dự báo lưu lượng dòng chảy”. Tuyển tập Hội nghị Khoa học thường niên năm 2018, Đại học Thủy lợi. Trang 560-562.
- [3] Hồ Việt Hùng, Lê Xuân Hiền, Giha Lee (2018). “Ứng dụng mạng thần kinh nhân tạo dự báo lưu lượng dòng chảy sông Hồng tại Sơn Tây dựa trên dữ liệu ở thượng lưu”. Hội nghị Khoa học thường niên năm 2018, Hội Cơ học Thủy khí.
- [4] Hồ Việt Hùng (2019). “Ứng dụng học máy (machine learning) trong việc dự báo lưu lượng dòng chảy và đỉnh lũ sông Hồng”. Tuyển tập Hội nghị Khoa học Cơ học Thủy khí toàn quốc lần thứ 22, trang 381-390.
- [5] Hồ Việt Tuấn, Hồ Việt Hùng (2019). “Sử dụng mạng nơ-ron nhân tạo dự báo mực nước sông

chịu ảnh hưởng của thủy triều”. Tạp chí Khoa học và công nghệ Thủy lợi, Viện Khoa học Thủy lợi Việt Nam, số 52, trang 108-116.

- [6] Xuan-Hien Le, Hung Viet Ho, Giha Lee, Sungho Jung (2019). “Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting”. *Water*, MDPI, 11, 1387; doi:10.3390/w11071387. <https://doi.org/10.3390/w11071387>
- [7] Asaad Y. Shamseldin (2010). “Artificial neural network model for river flow forecasting in a developing country”. *Journal of Hydroinformatics*, 12.1.
- [8] Chen, J.F., Hsieh, H.N., and Do, Q.H. (2014). “Forecasting Hoabinh Reservoir’s Incoming Flow: An Application of Neural Networks with the Cuckoo Search Algorithm”. *Information*. 5, 570-586.
- [9] Le, X.H., Ho, V.H., Lee, G.H., and Jung, S.H. (2018) “A Deep Neural Network Application for Forecasting the Inflow into the Hoabinh Reservoir in Vietnam”. *Proceeding of International Symposium on Lowland Technology*. pp. 151-159.
- [10] Sung, J.Y., Lee, J., Chung, I.M., and Heo, J.H. (2017). “Hourly water level forecasting at tributary affected by main river condition”. *KSCE Journal of Civil Engineering*. 9, pp. 644-653.
- [11] Veintimilla-Reyes J, Cisneros F, Vanegas P. (2016). “Artificial neural networks applied to flow prediction: A use case for the Tomebamba River”. *Procedia Engineering* 162: pp. 153-161.
- [12] Thông tư số 42/2017/TT-BTNMT, ngày 23/10/2017 Quy định kỹ thuật đánh giá chất lượng dự báo, cảnh báo thủy văn.