

# NGHIÊN CỨU ỨNG DỤNG MÔ HÌNH HỌC MÁY VÀO DỰ BÁO MỨC NƯỚC HỒ BUÔN TUA SARH

Trần Thị Tuyết, Đỗ Anh Đức, Hoàng Diệu Hằng

Viện Thủy điện và Năng lượng Tái tạo

Đặng Thanh Tuấn

Trung tâm Công nghệ phần mềm Thủy lợi

**Tóm tắt:** Nghiên cứu này áp dụng hai mô hình học máy là SVR và LSTM để dự báo mực nước hồ Buôn Tua Srah, lưu vực Srê Pôk, Việt Nam. Các bước dự báo khác nhau (6h, 12h, 24h, 48h, 72h) được thực hiện với các trường hợp dữ liệu đầu vào khác nhau được thu thập từ 2017-2023 ở thời điểm dự báo, 24h và 72h trước đây để tìm ra bộ dữ liệu đầu vào phù hợp cho mỗi mô hình học máy. Kết quả nghiên cứu cho thấy mô hình SVR và LSTM không có nhiều khác biệt trong dự báo mực nước 6h, 12h và 24h tới với sai số MAE thấp nhất tương ứng là 0,03, 0,05 và 0,08m. Đối với dự báo 48h và 72h, mô hình SVR cho kết quả dự báo tốt hơn so với mô hình LSTM. Bộ dữ liệu đầu vào cho mô hình SVR bao gồm mực nước hồ, dòng chảy đến, xả và thể tích hồ tại thời điểm 24h trước cho kết quả dự báo tốt nhất cho các bước 6h, 12h và 24h trong khi mô hình LSTM cho kết quả dự báo tốt hơn chỉ với dữ liệu mực nước hồ 72h trước đây.

**Từ khóa:** Dự báo mực nước hồ, mô hình học máy, SVR, LSTM, ...

**Summary:** This study applies two machine learning models, SVR and LSTM, to forecast the water levels of Buon Tua Srah reservoir, located in the Srepok basin, Vietnam. Various forecast intervals (6h, 12h, 24h, 48h, 72h) were conducted using different input data, collected from 2017 to 2023, at the forecast time, as well as 24h and 72h prior, to determine the most suitable input dataset for the models. The results indicate that there is little difference between the SVR and LSTM models in forecasting water levels 6h, 12h, and 24h ahead, with the lowest MAE errors being 0.03m, 0.05m, and 0.08m, respectively. For 48h and 72h forecasts, the SVR model provided better predictions compared to the LSTM model. The optimal input dataset for the SVR model included water level, inflow, outflow, and reservoir volume data from 24h prior, yielding the best results for the 6h, 12h, and 24h steps. In contrast, the LSTM model performed better with only the water level data from 72h prior.

**Keyword:** Reservoir Water Level; Prediction; Deep Learning

## 1. ĐẶT VẤN ĐỀ

Hồ chứa là công trình chứa nước trong lưu vực sông đóng vai trò quan trọng trong việc quản lý, quy hoạch và sử dụng tài nguyên nước. Hồ chứa được sử dụng cho các mục đích khác nhau như kiểm soát lũ lụt, sản xuất thủy điện, tưới tiêu trong nông nghiệp và giảm hạn hán [1]. Vì thế việc dự báo được mực nước hồ chính xác trong tương lai là yếu tố quan trọng

để vận hành an toàn hồ chứa cũng như việc tối ưu hóa chức năng của hồ.

Trong công tác dự báo mực nước, các phương pháp chủ yếu được chia thành hai loại: mô hình dựa trên vật lý (physics-based models) và phương pháp tiếp cận dựa trên dữ liệu (data-driven models). Phương pháp dựa trên quá trình vật lý thường mô tả đầy đủ bản chất vật lý của quá trình mưa – dòng chảy, thủy văn – thủy lực thông qua các phương trình toán học [2]. Mặc dù các mô hình vật lý cho thấy được khả năng ứng dụng cao trong việc tính toán mô phỏng các quá trình thủy văn, thủy

Ngày nhận bài: 25/6/2024

Ngày thông qua phản biện: 10/7/2024

Ngày duyệt đăng: 22/7/2024

lực, đưa ra dự báo nhiều kịch bản khác nhau, tuy nhiên việc sử dụng các mô hình này thường có khối lượng tính toán lớn và yêu cầu lượng dữ liệu chi tiết về các đặc tính của lưu vực như đặc điểm địa hình địa mạo, số liệu thông tin địa lý kết hợp các dữ liệu khác về mưa, dòng chảy, bốc hơi,... Các phương pháp tiếp cận dựa trên dữ liệu thường dựa trên việc phân tích các chuỗi dữ liệu đầu vào và đầu ra từ đó xác định mối quan hệ giữa mà chúng không đòi hỏi sự hiểu biết rõ về các quá trình vật lý đằng sau thông qua các phương pháp phân tích thống kê hay các phương pháp học máy (ML)[3]. Các mô hình thống kê thường được xây dựng dựa trên giả thiết rằng quá trình tạo dòng chảy tuân theo một phân phối chuẩn vì thế các mô hình này có độ chính xác hạn chế trong công tác dự báo do quá trình dòng chảy thường mang tính ngẫu nhiên và phi tuyến tính [4]. Trong khi đó các mô hình ML tập trung nhiều hơn vào việc khai thác mẫu và xu hướng trong dữ liệu để dự đoán mà ít quan tâm đến hiểu biết về mối quan hệ cụ thể giữa các biến thông qua việc sử dụng các phương pháp hộp đen để thiết lập mối quan hệ tối ưu giữa dữ liệu đầu vào và đầu ra [5]. Trong hai thập kỷ qua, cùng với sự phát triển của ngành khoa học máy tính, các mô hình ML đã chứng minh được những ưu điểm vượt trội so với mô hình vật lý và các phương pháp thống kê truyền thống với khả năng xử lý các mối quan hệ tương tác phức tạp và phi tuyến tính giữa các yếu tố hình thành dòng chảy với độ chính xác cao và khả năng xử lý, tính toán nhanh dữ liệu lớn một cách hiệu quả và tiết kiệm chi phí [6]. Hiện nay, bên cạnh những mô hình ML đã được ứng dụng từ rất lâu từ những năm 90 như ANN, LSTM (Long short-term memory) và SVM (Support Vector Machine) là hai mô hình được ứng dụng khá rộng rãi và hiệu quả trong các bài toán dự báo thủy văn.

Mô hình LSTM là mô hình học sâu – deep learning là một dạng của mô hình ML được thiết kế để xử lý và dự đoán chuỗi dữ liệu tuần tự (sequential data) với khả năng phân tích chuỗi thời gian và học từ dữ liệu không tuyến tính và phức tạp. Cấu trúc của mô hình LSTM

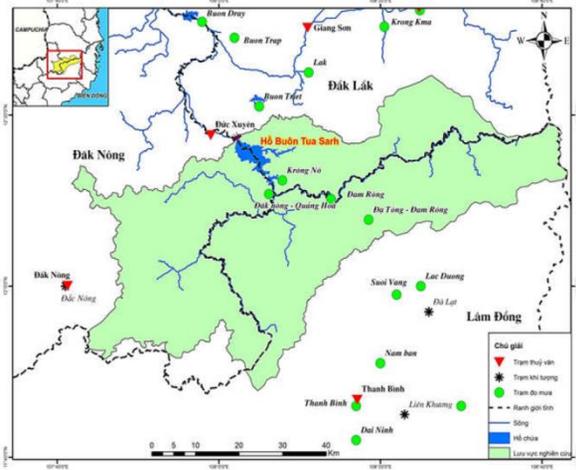
bao gồm các cổng (gates) giúp mô hình lưu trữ những thông tin quan trọng một cách có kiểm soát trong suốt quá trình huấn luyện, giúp cải thiện khả năng ghi nhớ và cho phép mô hình học từ các chuỗi dữ liệu dài mà không làm giảm hiệu suất của mô hình. Các nghiên cứu so sánh khả năng dự báo của mô hình LSTM so với các mô hình mạng nơ-ron truyền thống như RNN, ANN,... đều cho thấy LSTM cho kết quả dự báo tốt hơn các mô hình khác [7][8]. SVM là một phương pháp học máy có giám sát, có thể giải quyết các vấn đề về tuyến tính, phi tuyến tính và có khả năng tổng quát hóa mạnh mẽ do đó có thể ứng dụng tốt trong các bài toán về phân loại, hồi quy và dự báo chuỗi [9]. Mô hình SVM với các biến thể LS – SVR hay SVR cho thấy kết quả dự báo tốt hơn và khả năng dự báo tốt hơn khi so sánh với các phương pháp học máy khác[10]. Những ứng dụng phổ biến của mô hình trong các bài toán dự báo thủy văn khác nhau bao gồm dự báo mưa, dự báo dòng chảy, dự báo lũ,... [11]. Nghiên cứu của [12] đã sử dụng mô hình SVR để mô phỏng vận hành của hồ chứa thủy điện Gezhouba thuộc tỉnh Hồ Bắc, Trung Quốc theo các thời đoạn giờ, ngày, tháng, kết quả cho thấy có thể mô phỏng vận hành hồ chứa ngay cả với trường hợp dữ liệu đầu vào là rất ít. Tại Việt Nam, mô hình cũng được ứng dụng để dự báo dòng chảy cho lưu vực sông Ba [13] hay dự báo mực nước tại trạm Cao Lãnh, tỉnh Đồng Tháp [14],....

Như vậy với những tiềm năng trong ứng dụng dự báo thủy văn nói chung và mực nước nói riêng, nghiên cứu này được thực hiện nhằm mục đích áp dụng và so sánh khả năng dự báo mực nước của hai mô hình AI là SVR và LSTM, nghiên cứu tập trung vào phần thượng nguồn sông Srê Pôk với hồ thủy điện Buôn Tua Sarh, một trong những hồ chứa thủy điện lớn nhất của phần thượng nguồn sông Srepok. Thông qua việc ứng dụng hai mô hình trong dự báo mực nước, với việc phân tích hiệu quả dự báo của từng mô hình thông qua các chỉ số đánh giá, nghiên cứu nhằm mục đích lựa chọn

mô hình học máy phù hợp, đơn giản, dễ sử dụng và tính toán nhanh với dữ liệu đầu vào thích hợp, từ đó cung cấp cơ sở khoa học cho việc quản lý vận hành hồ chứa, cảnh báo lũ và phân bổ nguồn nước hợp lý.

## 2. DỮ LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

### 2.1 Giới thiệu khu vực nghiên cứu



Hình 1: Bản đồ khu vực nghiên cứu và mạng lưới trạm khí tượng, thủy văn

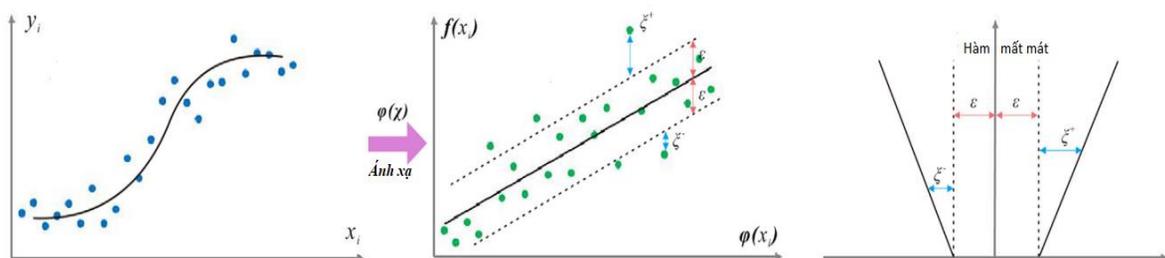
Sông Srêpôk là sông đa quốc gia thuộc sông Mê Công với phần thượng lưu nằm ở phía tây dãy Trường Sơn, thuộc lãnh thổ Việt Nam có diện tích khoảng 18.230 km<sup>2</sup>. Hồ thủy điện Buôn Tua Srah nằm trên địa bàn thôn Phú Vinh, xã Quảng Phú, huyện Krông Nô là hồ đầu nguồn quy mô lớn nhất trên lưu vực sông Srêpôk với diện tích lưu vực là 2930 km<sup>2</sup>. Hồ có dung tích toàn bộ là 789,6 triệu m<sup>3</sup>, lưu lượng trung bình nhiều năm đạt 102 m<sup>3</sup>/s.

Công trình thủy điện Buôn Tua Srah (Công trình cấp II) được xây dựng trên sông Krông Nô – nhánh chính của sông Srêpôk (huyện Lắk, tỉnh Đắk Lắk), chiều dài đỉnh đập chính 1,041m, chiều rộng đỉnh đập 8m. Hồ có nhiệm vụ cung cấp cho hệ thống điện quốc gia với công suất lắp máy 86MW, công suất đảm bảo 21MW, điện lượng trung bình hàng năm 358,5 triệu kWh. Việc vận hành an toàn, hợp lý của hồ có vai trò quan trọng hạn chế lũ cho vùng hạ du sông Srêpôk vào mùa mưa và tạo nguồn nước tưới, giảm thiểu hạn hán vào mùa kiệt, vì vậy nghiên cứu đã lựa chọn hồ thủy điện Buôn Tua Srah để phát triển mô hình.

### 2.2. Phương pháp nghiên cứu

#### 2.1.1 Mô hình Support Vector Regression

SVM là một thuật toán học máy có giám sát được đề xuất đầu tiên bởi Vladimir N. Vapnik vào năm 1963 [15]. Thuật toán SVM được xây dựng dựa trên ý tưởng tìm kiếm siêu phẳng (hyperplane) để phân tách các điểm dữ liệu. Siêu phẳng thường được tối ưu thông qua việc sử dụng thủ thuật để ánh xạ tập dữ liệu ban đầu (kernel trick) lên không gian nhiều chiều hơn, ở đó việc tìm kiếm siêu phẳng tối ưu được chứng minh là đơn giản hơn [16]. SVM có hàm mục tiêu là tối đa hóa biên/lề (margin) – khoảng cách giữa siêu phẳng đến các điểm dữ liệu gần nhất tương ứng với các phân lớp, và các điểm trên được gọi là support vectors. SVM được thay đổi để áp dụng vào các bài toán hồi quy với tên gọi SVR (Support Vector Regression).



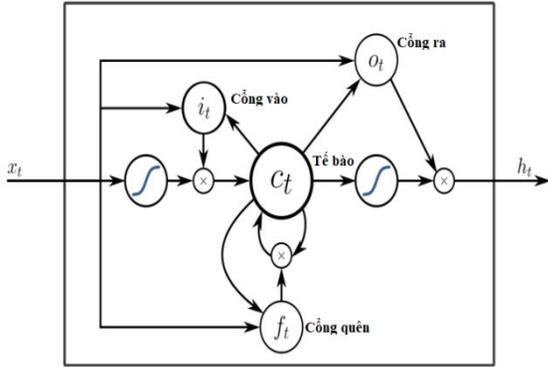
Hình 2: Minh họa khái quát SVR

Với tập dữ liệu huấn luyện là  $\{X_i, Y_i\}_{i=1}^I$ , trong đó  $I$  là số lượng điểm dữ liệu, hàm  $f(x)$

miêu tả mối quan hệ phi tuyến giữa biến  $x_i$  và  $y_i$  như sau:

$$f(x) = \sum_{i=1}^I (a_i - a_i^*) K(x, x_i) + b \quad (1)$$

Trong đó  $a_i$  và  $a_i^*$  là các nhân tử Lagrange,  $K$  là hàm nhân (kernel function),  $b$  là hệ số thiên lệch. Các hàm nhân phổ biến là Linear, Polynomial, Gaussian và Sigmoid.



Hình 3: Cấu trúc của một LSTM neuron

### 2.1.2 Mô hình Long Short-Term Memory

Mô hình LSTM là một dạng mở rộng của mô hình mạng nơron hồi quy ((Recurrent Neural Networks -RNN) [17] đã được đề xuất bởi Hochreiter & Schmidhuber (1997)[18] được cho là tối ưu hơn phương pháp mạng nơron truyền thống khi cho phép RNN ghi nhớ đầu vào của chúng trong một thời gian dài. LSTM thường được sử dụng rộng rãi trong các bài toán liên quan đến dự báo dựa vào dữ liệu chuỗi thời gian. Cấu trúc của mô hình LSTM bao gồm ba cổng (hình 3): Cổng vào (input gate), cổng quên (forget gate) và cổng ra (output gate). Cổng quên (forget gate) sẽ quyết định thông tin nào cần loại bỏ từ tế bào. Cổng vào (input gate) sẽ quyết định xem thông tin mới nào sẽ được lưu vào trạng thái tế bào. Cổng đầu ra (output gate) trích xuất thông tin hữu ích từ trạng thái hiện tại để trình bày ở định dạng đầu ra. Các cổng được kết hợp bởi một tầng mạng sigmoid và một phép nhân để sàng lọc thông tin đi qua nó

### 2.3. Dữ liệu đầu vào

Dữ liệu được thu thập, xử lý dữ liệu thiếu hoặc bị nhiễu sau đó được chuẩn hóa và đưa vào mô hình bao gồm các dữ liệu khí tượng, thủy văn và dữ liệu hồ chứa.

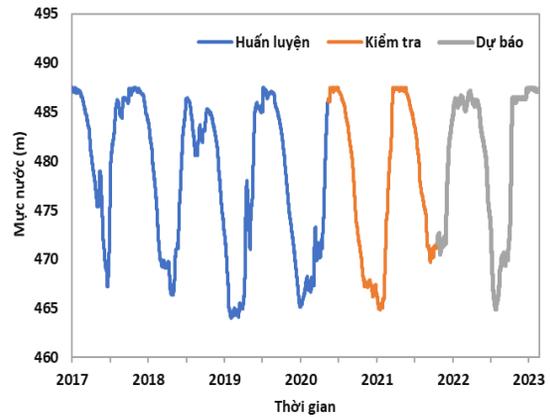
#### 2.1.1 Dữ liệu khí tượng, thủy văn

- Dữ liệu mưa thực đo 1 giờ từ năm 2017 - 2023 của các trạm đo mưa Krong No và trạm Đắc Nông trong lưu vực nghiên cứu.

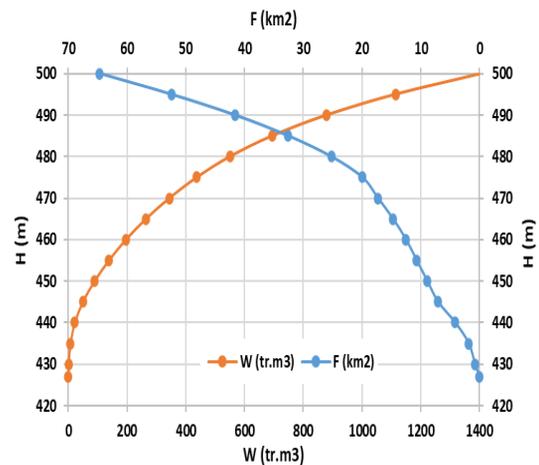
- Dữ liệu về lưu lượng nước vào hồ và ra khỏi hồ cũng được thu thập và xử lý theo bước thời gian 1 giờ từ năm 2017-2023

#### 2.1.2 Dữ liệu hồ chứa

- Dữ liệu mực nước hồ theo bước thời gian 1h từ năm 2017-2023 (hình 4).



Hình 4: Dữ liệu mực nước hồ (2017-2023)



Hình 5: Đặc tính lòng hồ Buôn Tua Sarh

Ngoài ra sự thay đổi mực nước hồ còn liên quan đến quan hệ dung tích hồ và diện tích mặt nước của hồ, vì thế nghiên cứu đã sử dụng quan hệ Z-F-V (hình 5) để từ đó nội suy các giá trị thể tích hồ chứa và diện tích mặt nước của hồ theo thời gian tương ứng với mực nước hồ.

### 2.4. Thiết lập mô hình

Chuỗi dữ liệu được chia thành 3 phần: 70%

cho tập huấn luyện (training) và 30% cho tập kiểm tra (testing) và còn lại là tập dự báo thử nghiệm (hình 4b). Trong quá trình huấn luyện các thông số được lựa chọn dựa vào các chỉ số của mô hình theo các trường hợp dữ liệu khác nhau. Các mô hình ML luôn hoạt động theo một nguyên tắc đó là học từ tất cả các dữ liệu được đưa vào vì vậy lựa chọn dữ liệu đầu vào ảnh hưởng trực tiếp đến việc cải thiện hiệu quả dự báo của mô hình, tăng tốc độ tính toán và để hiểu rõ hơn các quá trình ẩn đằng sau. Do vậy nghiên cứu tập trung vào việc khai thác các tập dữ liệu khác nhau ảnh hưởng đến hoạt động của hồ chứa kết hợp với bước thời gian khác nhau để tìm bộ dữ liệu thích hợp cho việc dự báo mực nước hồ. Để dự báo mực nước hồ tại các thời điểm t+6, t+12 và t+24, t+48, t+72, nghiên cứu sử dụng các trường hợp dữ liệu khác nhau (4 trường hợp) tương ứng ở những bước thời gian dữ liệu trong quá khứ khác nhau (3 kịch bản).

Các trường hợp dữ liệu đầu vào khác nhau bao gồm:

- TH1: Dữ liệu mưa, dữ liệu mực nước hồ, lưu lượng đến hồ, lưu lượng ra khỏi hồ, diện tích và dung tích hồ (X, H<sub>h</sub>, Q<sub>đ</sub>, Q<sub>r</sub>, F<sub>h</sub>, V<sub>h</sub>)
- TH2: Dữ liệu mực nước hồ, lưu lượng đến hồ, lưu lượng ra khỏi hồ, diện tích và dung tích hồ (H<sub>h</sub>, Q<sub>đ</sub>, Q<sub>r</sub>, F<sub>h</sub>, V<sub>h</sub>)
- TH3: Dữ liệu mực nước hồ, lưu lượng đến hồ, lưu lượng ra khỏi hồ (H<sub>h</sub>, Q<sub>đ</sub>, Q<sub>r</sub>)
- TH4: Dữ liệu mực nước hồ (H<sub>h</sub>)

tương ứng dữ liệu trong quá khứ với KB1 là dữ liệu tại bước thời gian (t), KB2 là dữ liệu tại bước thời gian t, t-1,...t-24 và KB3 là dữ liệu tại bước thời gian t, t-1,...t-72.

Đối với mô hình SVR, để tránh tình trạng overfit của mô hình, nghiên cứu đã sử dụng phương pháp kiểm định chéo nhiều lớp (k-fold cross validation) do tính đơn giản và hiệu quả cao trong việc sử dụng. Phương pháp tối ưu GridSearchCV cũng được sử dụng trong mô hình SVR để tìm được bộ thông số tối ưu.

### 2.5. Phương pháp đánh giá mô hình

Để đánh giá hiệu quả dự báo của các mô hình, nghiên cứu này đã sử dụng các chỉ số thống kê thường dùng trong đánh giá mực nước bao gồm chỉ số RMSE - Root Mean Square Error, sai số trung bình tuyệt đối Mean Absolute Error (MAE) và hệ số tương quan R<sup>2</sup>

RMSE là căn bậc hai của trung bình các bình phương sai số và thường nhạy cảm với các sai số lớn. Công thức tính toán chỉ số RMSE:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i^{obs} - Y_i^{pre})^2}{n}} \quad (2)$$

Trong đó  $Y_i^{obs}$ ,  $Y_i^{pre}$  tương ứng lần lượt là giá trị thực tế và giá trị dự báo, n là tổng số giá trị thực đo.

MAE là giá trị trung bình tuyệt đối của các sai số giữa giá trị dự báo và giá trị thực đo

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i^{obs} - Y_i^{pre}| \quad (3)$$

Hệ số tương quan R<sup>2</sup> được tính theo công thức sau:

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i^{pre} - \bar{Y}^{pre})^2}{\sum_{i=1}^n (Y_i^{obs} - \bar{Y}^{obs})^2} \quad (4)$$

Trong đó  $\bar{Y}_i^{pre}$  và  $\bar{Y}_i^{obs}$  lần lượt là giá trị trung bình dự báo và thực đo

### 3. KẾT QUẢ NGHIÊN CỨU

**Bảng 1: Kết quả đánh giá khả năng dự báo mực nước của mô hình LSTM**

LSTM		KB1				KB2				KB3			
		TH1	TH2	TH3	TH4	TH1	TH2	TH3	TH4	TH1	TH2	TH3	TH4
MAE	6h	0,03	0,03	0,03	0,05	0,05	0,03	0,03	0,03	0,13	0,05	0,04	0,04
	12h	0,06	0,07	0,06	0,06	0,07	0,05	0,05	0,05	0,15	0,10	0,11	0,06
	24h	0,12	0,12	0,10	0,18	0,20	0,10	0,10	0,09	0,45	0,16	0,15	0,13
	48h	0,30	0,27	0,25	0,28	0,38	0,32	0,26	0,17	0,46	0,48	0,30	0,18
	72h	0,43	0,42	0,39	0,41	0,59	0,51	0,47	0,25	0,73	0,44	0,48	0,29

LSTM		KB1				KB2				KB3			
		TH1	TH2	TH3	TH4	TH1	TH2	TH3	TH4	TH1	TH2	TH3	TH4
RMSE	6h	0,05	0,05	0,04	0,07	0,06	0,04	0,04	0,04	0,16	0,09	0,07	0,05
	12h	0,08	0,11	0,08	0,08	0,09	0,07	0,07	0,07	0,22	0,13	0,16	0,08
	24h	0,18	0,20	0,15	0,28	0,27	0,15	0,17	0,13	0,94	0,25	0,25	0,18
	48h	0,45	0,39	0,40	0,54	0,56	0,44	0,37	0,26	0,87	0,73	0,46	0,29
	72h	0,65	0,61	0,59	0,80	0,85	0,71	0,67	0,41	1,26	0,74	0,81	0,46
R <sup>2</sup>	6h	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	12h	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	24h	1,000	1,000	1,000	0,999	0,999	1,000	1,000	1,000	0,993	1,000	1,000	1,000
	48h	0,998	0,999	0,999	0,998	0,998	0,999	0,999	0,999	0,994	0,996	0,998	0,999
	72h	0,997	0,997	0,998	0,996	0,996	0,997	0,997	0,999	0,992	0,996	0,995	0,998

**Bảng 2: Kết quả đánh giá khả năng dự báo mực nước của mô hình SVR**

SVR		KB1				KB2				KB3			
		TH1	TH2	TH3	TH4	TH1	TH2	TH3	TH4	TH1	TH2	TH3	TH4
MAE	6h	0,03	0,03	0,03	0,05	0,04	0,03	0,03	0,03	0,05	0,04	0,04	0,03
	12h	0,06	0,06	0,06	0,09	0,05	0,05	0,05	0,05	0,07	0,06	0,06	0,05
	24h	0,11	0,10	0,10	0,10	0,09	0,08	0,09	0,08	0,11	0,08	0,10	0,08
	48h	0,20	0,21	0,21	0,29	0,20	0,21	0,21	0,18	0,25	0,19	0,19	0,17
	72h	0,30	0,31	0,31	0,41	0,28	0,31	0,31	0,28	0,33	0,27	0,28	0,26
RMSE	6h	0,04	0,04	0,04	0,07	0,04	0,04	0,04	0,05	0,13	0,05	0,05	0,05
	12h	0,08	0,08	0,08	0,15	0,08	0,07	0,06	0,07	0,15	0,11	0,11	0,07
	24h	0,16	0,15	0,15	0,15	0,15	0,12	0,12	0,13	0,21	0,12	0,18	0,12
	48h	0,30	0,31	0,31	0,54	0,34	0,31	0,31	0,29	0,43	0,30	0,31	0,28
	72h	0,46	0,47	0,47	0,80	0,47	0,47	0,47	0,48	0,52	0,45	0,46	0,44
R <sup>2</sup>	6h	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	12h	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	24h	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	48h	0,999	0,999	0,999	0,998	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999
	72h	0,998	0,998	0,998	0,995	0,998	0,998	0,998	0,998	0,998	0,998	0,998	0,999

Kết quả tính toán các chỉ số dự báo của hai mô hình tương ứng ở Bảng 1 và 2 cho thấy sai số tuyệt đối trung bình MAE và sai số RMSE tăng dần khi thời gian dự báo tăng đồng thời hệ số tương quan của kết quả dự báo ở các trường hợp tính toán cũng giảm dần. Có thể thấy đối với dự báo 6h và 12h cả hai mô hình dự báo tương đối tốt với các trường hợp dữ liệu đầu vào khác nhau. Dự báo 6h cho kết quả dự báo trong khoảng từ 0,03- 0,05 m trừ kết quả dự báo từ mô hình LSTM với trường hợp dữ liệu đầu vào ở KB3(TH1) với sai số lên tới 0,13 m. Dự báo 12h cho kết quả dự báo trong khoảng 0,05-0,07 m, riêng với mô hình LSTM sai số MAE > 0,1m với bước thời gian sử dụng

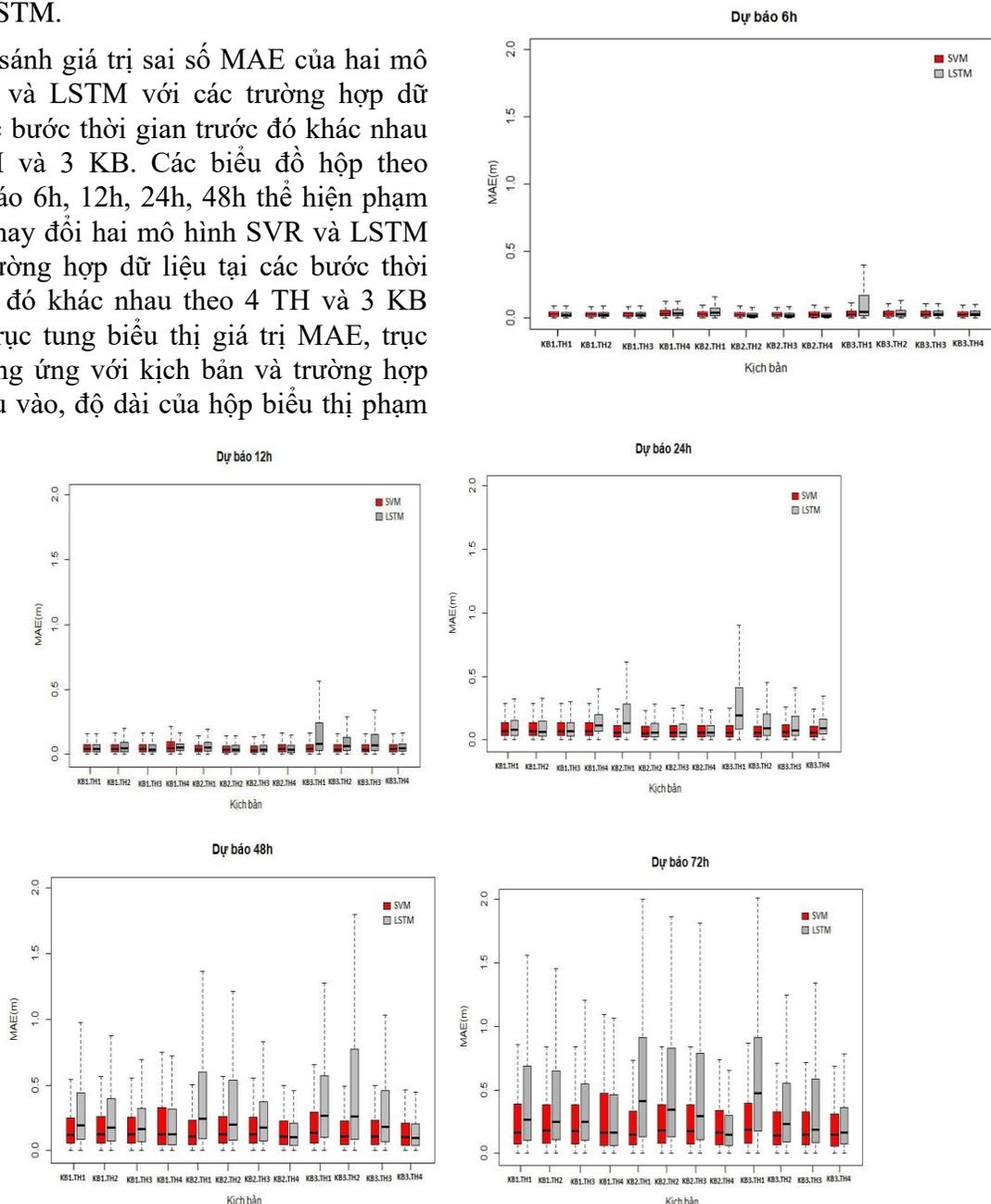
dữ liệu 72h trước đây (KB3) tương ứng với các trường dữ liệu TH1, TH2, TH3. Ở các trường hợp dự báo xa hơn 24h, 48h và 72h mô hình SVR cho kết quả với sai số dự báo thấp hơn mô hình LSTM.

- Kết quả dự báo theo chỉ số từ bảng 1 cũng cho thấy mô hình LSTM cho dự báo tốt nhất ở tất cả các bước thời gian dự báo với trường hợp chỉ sử dụng dữ liệu mực nước hồ trong quá khứ ở 24 h trước đây, sai số lớn nhất với trường hợp sử dụng tất cả các dữ liệu liên quan ở 72h trước đây (KB3.TH1). Đối với mô hình SVR, kết quả dự báo tốt nhất đối với dự báo 6h, 12h, 24h tương ứng với trường hợp dữ liệu

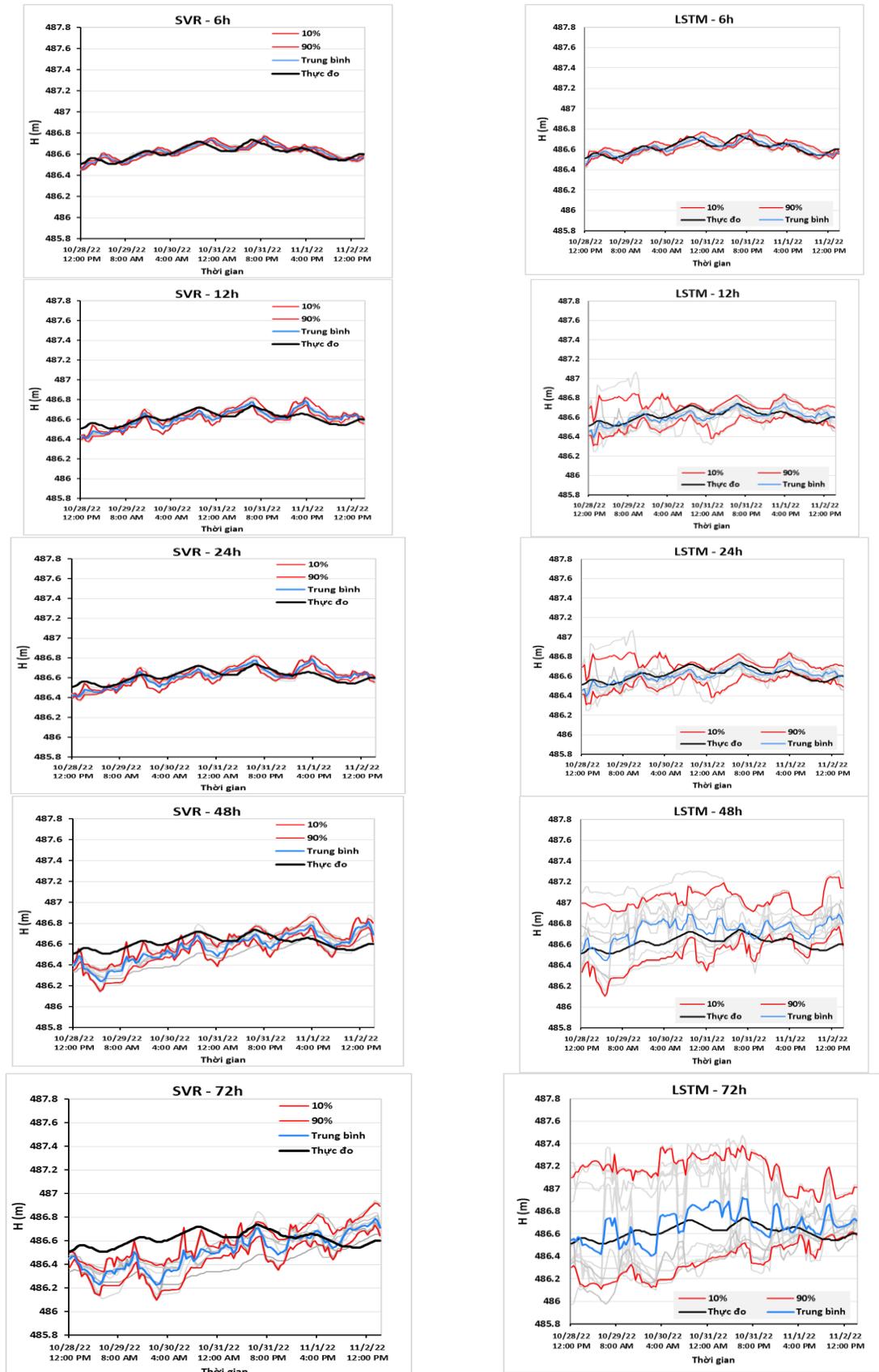
đầu vào là TH2 (  $H_{hồ}$ ,  $Q_{đến}$ ,  $Q_{xa}$ ,  $V_{hồ}$  ) ở 24h trước đấy (KB2) với sai số MAE tương ứng là 0,03, 0,05 và 0,08m. Với thời gian dự báo dài hơn, mô hình SVR cho kết quả dự báo tốt hơn khi sử dụng dữ liệu mực nước hồ ở 72h trước đấy (KB3.TH4). Như vậy có thể thấy việc sử dụng dữ liệu dài hơn trong quá khứ giúp cải thiện dự báo của mô hình SVR ở dự báo 48h và 72 h tới nhưng không cải thiện dự báo của mô hình LSTM.

Hình 5 so sánh giá trị sai số MAE của hai mô hình SVR và LSTM với các trường hợp dữ liệu tại các bước thời gian trước đó khác nhau theo 4 TH và 3 KB. Các biểu đồ hộp theo bước dự báo 6h, 12h, 24h, 48h thể hiện phạm vi giá trị thay đổi hai mô hình SVR và LSTM với các trường hợp dữ liệu tại các bước thời gian trước đó khác nhau theo 4 TH và 3 KB trong đó trục tung biểu thị giá trị MAE, trục hoành tương ứng với kịch bản và trường hợp dữ liệu đầu vào, độ dài của hộp biểu thị phạm

vi của 50% dữ liệu trung tâm tương ứng với khoảng phân vị 25% và 75%. Nhìn vào hình 5 cũng cho thấy khoảng phân tán của mô hình LSTM tương đối rộng và nghiêng về phía sai số lớn, đặc biệt ở dự báo 48h và 72h. Kết quả mô hình SVR cũng cho thấy khoảng phân vị của dự báo 48h và 72h rộng hơn so với dự báo 6h,12h và 24h, tuy nhiên thấp hơn so với kết quả từ mô hình LSTM.



Hình 5: So sánh kết quả dự báo của hai mô hình SVR và LSTM theo chỉ số MAE



Hình 6: Kết quả dự báo từ ngày 28/10/2022 00:00- 2/11/2022 15:00 của hai mô hình tương ứng với thời gian dự báo 6h,12h,24h,48h,72h

Kết quả dự báo trận lũ được trích xuất từ ngày 28/10/2022 00:00- 2/11/2022 15:00 theo hình 6, đường màu đen thể hiện mực nước hồ thực đo theo giờ, đường màu đỏ thể hiện khoảng giá trị từ 10-90% kết quả các kịch bản của hai mô hình, đường màu xanh thể hiện mực nước trung bình từ các kết quả dự báo. Kết quả cho thấy dự báo từ mô hình LSTM có sự sai khác rất lớn đối với các trường hợp dữ liệu đầu vào khác nhau, dự báo càng xa sự chênh lệch giữa khoảng phân vị 10% và 90 % càng lớn, kết quả giá trị trung bình của các kịch bản tính toán từ mô hình xoay quanh giá trị mực nước hồ thực tế. Kết quả dự báo từ mô hình SVR có khoảng phân vị 10-90% nhỏ hơn so với mô hình LSTM.

#### 4. KẾT LUẬN

Nghiên cứu đã bước đầu thử nghiệm thành công hai mô hình học máy là SVR và LSTM trong dự báo mực nước hồ Buôn Tua Srah với các trường dữ liệu đầu vào sẵn có, dễ sử dụng giúp hỗ trợ ra quyết định nhanh chóng trong công tác dự quản lý và vận hành hồ chứa. Các bước thời gian dự báo khác nhau bao gồm 6h, 12h, 24h, 48h, 72h sử dụng các trường hợp dữ liệu đầu vào khác nhau từ mực nước hồ, dòng chảy đến hồ, lượng nước ra khỏi hồ, quan hệ Z-V-F hồ và lượng mưa tại tác thời đoạn t, 24h và 72 h trước để đánh giá và đề xuất lựa chọn bộ dữ liệu đầu vào phù hợp cho mô hình dự báo mực nước hồ. Kết quả nghiên cứu cho thấy mô hình SVR và LSTM không có nhiều khác biệt khi dự báo mực nước hồ 6h và 12h tới, tuy nhiên mô hình SVR ổn định hơn khi dự đoán mực nước hồ ở cái thời điểm xa hơn cho thấy khả năng khái quát hóa tương đối tốt của mô hình. Bộ dữ liệu đầu vào ở TH2 bao gồm  $H_{hồ}$ ,  $Q_{đến}$ ,  $Q_{xa}$ ,  $V_{hồ}$  ở thời điểm 24 h trước đây cũng cho kết quả dự báo tốt hơn ở các thời điểm dự báo 6h, 12h và 24h. Với dự báo 48h và 72 h việc sử dụng dữ liệu ở 72 giờ

trước đây giúp cải thiện khả năng dự báo của mô hình. Đối với mô hình LSTM, dữ liệu đầu vào là mực nước 72h trước đây cho kết quả dự báo tốt nhất với các bước thời gian dự báo khác nhau. Có thể thấy sự khác biệt trong việc sử dụng bộ dữ liệu đầu vào khác nhau, ảnh hưởng đến kết quả dự báo của hai mô hình.

Những sai số trong quá trình ứng dụng hai mô hình học máy SVR và LSTM trong quá trình dự báo mực nước hồ có thể đến từ việc yêu cầu dữ liệu lớn để huấn luyện mô hình một cách hiệu quả, từ đó mô hình có khả năng khái quát tốt hơn. Mặc dù nghiên cứu đã áp dụng các phương pháp hạn chế hiện tượng quá khớp (overfitting), tuy nhiên những sai sót trong quá trình xử lý dữ liệu cũng có thể ảnh hưởng đến chất lượng dự báo của mô hình. Kết quả dự báo từ mô hình LSTM trong trường hợp sử dụng đầy đủ các dữ liệu liên quan bao gồm cả dữ liệu mưa có độ chính xác thấp hơn đáng kể so với các trường hợp dữ liệu khác có thể bắt nguồn từ việc những dao động trong dữ liệu này cao hơn so với dữ liệu mực nước hồ.

Dựa trên những phân tích và đánh giá đã thực hiện, nhóm nghiên cứu đề xuất sử dụng mô hình SVR cho dự báo mực nước hồ chứa Buôn Tua Sarh 6h, 12h và 24 h tới với bộ dữ liệu đầu vào bao gồm mực nước, lưu lượng đến và ra khỏi hồ, diện tích mặt nước và dung tích hồ ở 24h trước đây. Trong trường hợp dữ liệu hạn chế hơn, chỉ có dữ liệu mực nước hồ trong quá khứ, việc áp dụng mô hình LSTM cũng cho kết quả dự báo tốt với dữ liệu mực nước 24 giờ trước đây. Đối với những dự báo dài hơn, cần nghiên cứu khai thác tối đa thông tin từ các dữ liệu khí tượng thủy văn, giúp kéo dài chuỗi giúp tăng khả năng “học” và khái quát của mô hình, ngoài ra có thể sử dụng các mô hình lai (hybrid models) cũng giúp cải thiện độ chính xác và khả năng tổng quát hóa của dự báo.

#### TÀI LIỆU THAM KHẢO

- [1] M. Liu *et al.*, “The applicability of lstm-knn model for real-time flood forecasting in different climate zones in China,” *Water (Switzerland)*, vol. 12, no. 2, 2020.
- [2] C. Paniconi and M. Putti, “Physically based modeling in catchment hydrology at 50: Survey and outlook,” *Water Resour. Res.*, vol. 51, no. 9, pp. 7090–7129, 2015

- [3] D. P. Solomatine and A. Ostfeld, “Data-driven modelling: Some past experiences and new approaches,” *J. Hydroinformatics*, vol. 10, no. 1, pp. 3–22, 2008, doi: 10.2166/hydro.2008.015.
- [4] S. Ghimire, Z. M. Yaseen, A. A. Farooque, R. C. Deo, J. Zhang, and X. Tao, “Streamflow prediction using an integrated methodology based on convolutional neural network and long short-term memory networks,” *Sci. Rep.*, vol. 11, no. 1, pp. 1–27, 2021, doi: 10.1038/s41598-021-96751-4.
- [5] C. P. Van, H. D. Nguyen, Q. H. Nguyen, and Q. T. Bui, “Daily streamflow prediction based on the long short-term memory algorithm: a case study in the Vietnamese Mekong Delta,” *J. Water Clim. Chang.*, vol. 14, no. 4, pp. 1247–1267, 2023, doi: 10.2166/wcc.2023.419.
- [6] D. Zhang, Q. Peng, J. Lin, D. Wang, X. Liu, and J. Zhuang, “Simulating reservoir operation using a recurrent neural network algorithm,” *Water (Switzerland)*, vol. 11, no. 4, 2019, doi: 10.3390/w11040865.
- [7] C. Choi, J. Kim, H. Han, D. Han, and H. S. Kim, “Development of Water Level Prediction Models Using Machine Learning in Wetlands: A Case Study of Upo Wetland in South Korea,” *Water*, vol. 12, no. 93, 2020, doi: 10.3390/w12010093.
- [8] Z. Vizi *et al.*, “Water level prediction using long short-term memory neural network model for a lowland river: a case study on the Tisza River, Central Europe,” *Environ. Sci. Eur.*, vol. 35, no. 1, 2023, doi: 10.1186/s12302-023-00796-3.
- [9] K.-R. Müller, A. J. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen, and V. Vapnik, “Using support vector machines for time series prediction.” pp. 243–253, 1999.
- [10] O. Kisi and K. S. Parmar, “Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution,” *J. Hydrol.*, vol. 534, pp. 104–112, 2016.
- [11] J. Y. Lin, C. T. Cheng, and K. W. Chau, “Using support vector machines for long-term discharge prediction,” *Hydrol. Sci. J.*, vol. 51, no. 4, pp. 599–612, 2006, doi: 10.1623/hysj.51.4.599.
- [12] D. Zhang *et al.*, “Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm,” *J. Hydrol.*, vol. 565, pp. 720–736, 2018.
- [13] C. H. Hải, T. A. Phương, T. Q. Như, and T. M. Cường, “Áp dụng trí tuệ nhân tạo vào dự báo lưu lượng đến hồ lưu vực sông Ba,” *Tạp chí Khí tượng Thủy văn*, vol. 09, pp. 22–33, 2019.
- [14] L. Xuân Hòa and N. Tiền Giang, “Xây dựng các mô hình hồi quy hỗ trợ véc tơ dự báo mực nước trạm Cao Lãnh, tỉnh Đồng Tháp,” *Vietnam J. Hydrometeorol.*, vol. 8, no. 740(1), pp. 87–97, 2022, doi: 10.36335/vnjhm.2022(740(1)).87-97.
- [15] Vapnik and V. N., “The Nature of Statistical Learning,” *Theory*. p. 334, 1995.
- [16] B. E. Boser, I. M. Guyon, and V. N. Vapnik, “Training algorithm for optimal margin classifiers,” *Proc. Fifth Annu. ACM Work. Comput. Learn. Theory*, no. August 1996, pp. 144–152, 1992, doi: 10.1145/130385.130401.
- [17] F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, “Learning precise timing with LSTM recurrent networks,” *J. Mach. Learn. Res.*, vol. 3, no. 1, pp. 115–143, 2003, doi: 10.1162/153244303768966139.
- [18] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.