Tạp chí Khoa học Giao thông vận tải, Tập 76, Số 02 (02/2025), 172-186

Transport and Communications Science Journal

APPLICATION OF MACHINE LEARNING ALGORITHMS TO DETERMINE THE SEQUENCE DEPTH OF A HYDRAULIC JUMP IN A RECTANGULAR CHANNEL CONSIDERING THE EFFECT OF FRICTIONAL FORCE

Hung Viet Ho*

Faculty of Water Resources Engineering, Thuyloi University, 175 Tay Son, Hanoi, Viet Nam

ARTICLE INFO

TYPE: Research Article
Received: 21/11/2024
Revised: 07/02/2025
Accepted: 10/02/2025
Published online: 15/02/2025
https://doi.org/10.47869/tcsj.76.2.5

* *Corresponding author*
Email: hohung.thuyluc@tlu.edu.vn

Abstract. The sequence depth of a hydraulic jump is a critical component that greatly influences the depth and length of the energy dissipator. Calculating this depth correctly is vital and has important practical implications. As a result, this work established and assessed the predicting power of six machine learning (ML) models, including Random Forest (RT), Adaptive Boosting (Ada), Cat Boosting (CB), Gradient Boosting (GB), Extra Trees (ET), and Support Vector Machine (SVM), for sequence depth. The π-Buckingham theorem was applied in this study to identify five dimensionless parameters for the model's input and output. The research results show that the ET, GB, SVR, and Ada models, which consider the effects of roughness and channel width, as well as the viscosity of the fluid, have smaller errors compared to the Belanger formula (which ignores friction) and other empirical formulas. When evaluated, all these models had Nash coefficients higher than 0.996. The ET model produces the best outcomes, followed by the GB, SVR, Ada, RF, and CB models in decreasing order. Thus, the ET model can be used to calculate the hydraulic jump's sequence depth in a prismatic horizontal channel with a rectangular cross section.

Keywords: hydraulic jump, depth, machine learning, model, Froude

@ 2025 University of Transport and Communications

Transport and Communications Science Journal, Vol 76, Issue (02/2025), 172-186

Tạp chí Khoa học Giao thông vận tải

ỨNG DỤNG CÁC THUẬT TOÁN HỌC MÁY XÁC ĐỊNH ĐỘ SÂU SAU NƯỚC NHẢY TRONG KÊNH CHỮ NHẬT CÓ XÉT ĐẾN ẢNH HƯỞNG CỦA LỰC MA SÁT

Hồ Việt Hùng*

Khoa Kỹ thuật tài nguyên nước, Trường Đại học Thủy lợi, 175 Tây Sơn, Hà Nội, Việt Nam

THÔNG TIN BÀI BÁO

CHUYÊN MỤC: Công trình khoa học
Ngày nhận bài: 21/11/2024
Ngày nhận bài sửa: 07/02/2025
Ngày chấp nhận đăng: 10/02/2025
Ngày xuất bản Online: 15/02/2025
https://doi.org/10.47869/tcsj.76.2.5
* *Tác giả liên hệ*
Email: hohung.thuyluc@tlu.edu.vn

Tóm tắt. Độ sâu sau nước nhảy là một yếu tố quan trọng, ảnh hưởng rất lớn đến độ sâu và chiều dài của bể tiêu năng. Việc tính toán chính xác độ sâu này là rất cần thiết và có ý nghĩa thực tiễn cao. Do đó, nghiên cứu này đã thiết lập và đánh giá khả năng dự báo độ sâu sau nước nhảy của sáu mô hình học máy (ML), gồm có: Rừng cây ngẫu nhiên (Random Forest - RT), Tăng cường thích ứng (Adaptive Boosting – Ada), Tăng cường tốc độ (Cat Boosting – CB), Tăng cường độ dốc (Gradient Boosting - GB), Cây bổ sung (Extra Trees - ET) và Máy Vector hỗ trợ (Support Vector Machine – SVM). Trong nghiên cứu này, định lý π-Buckingham đã được sử dụng để tìm năm tham số không thứ nguyên làm đầu vào và đầu ra của mô hình. Kết quả nghiên cứu cho thấy, các mô hình ET, GB, SVR, Ada có xét đến ảnh hưởng của độ nhám và chiều rộng lòng dẫn, tính nhớt của chất lỏng, có sai số nhỏ hơn so với công thức Belanger (bỏ qua lực ma sát) và các công thức kinh nghiệm khác. Khi kiểm định, các mô hình này đều có hệ số Nash đạt trên 0,996. Mô hình ET cho kết quả tốt nhất, sau đó là GB, SVR, Ada, RF, CB, theo thứ tự giảm dần. Như vậy, có thể áp dụng mô hình ET để tính toán độ sâu sau nước nhảy trong kênh lăng trụ đáy bằng, mặt cắt chữ nhật.

Từ khóa: nước nhảy, độ sâu, học máy, mô hình, Froude

@ 2025 Trường Đại học Giao thông vận tải

1. GIỚI THIỆU CHUNG

Nước nhảy xảy ra khi dòng chảy chuyển từ trạng thái chảy xiết sang chảy êm ở sau đập tràn, cửa cống lộ thiên, hoặc dòng chảy xiết cuối cống ngầm hay trong kênh dẫn hở. Theo chiều dòng chảy, vận tốc và số Froude giảm đột ngột, độ sâu tăng đột ngột [1]. Các đặc trưng

hình học của nước nhảy cần được xác định gồm có: độ sâu trước và sau nước nhảy, chiều dài khu xoáy và chiều dài nước nhảy [2]. Độ sâu sau nước nhảy là một đặc trưng quan trọng của nước nhảy, ảnh hưởng trực tiếp đến độ sâu và chiều dài của bể tiêu năng. Do đó, việc tính toán chính xác độ sâu sau nước nhảy là rất cần thiết và có ý nghĩa thực tiễn cao.

Từ trước đến nay, độ sâu sau nước nhảy trong kênh chữ nhật nằm ngang thường được tính theo công thức Belanger (1828) khi biết số Froude và độ sâu trước nước nhảy [1,3]. Công thức này được thiết lập dựa trên Phương trình động lượng, khi bỏ qua lực ma sát và coi áp suất phân bố theo quy luật thủy tĩnh trong các mặt cắt trước và sau nước nhảy. Tỷ số hai độ sâu liên hiệp của nước nhảy (xem Hình 1) được tính theo công thức Belanger (1) như sau.

Belanger (1828):
$$Y = \frac{h_2}{h_1} = \frac{1}{2}\left(-1+\sqrt{1+8Fr_1^2}\right) \quad (1)$$

$$Fr_1 = \frac{V_1}{\sqrt{gh_1}} \quad (2)$$

Trong đó: h_1 và h_2 – các độ sâu trước và sau nước nhảy; Fr_1 – số Froude trước nước nhảy; V_1 – vận tốc trung bình tại mặt cắt trước nước nhảy; g – gia tốc trọng trường.

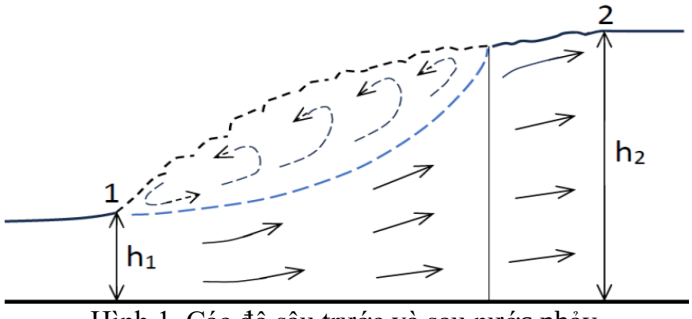
Nhiều nhà khoa học đã tiến hành nghiên cứu nước nhảy và đề xuất các công thức tính tỷ số hai độ sâu liên hiệp của nước nhảy theo số Froude trước nước nhảy. Sarma và Newnham (1973), đã chọn giá trị của hệ số sửa chữa động năng bằng 1,045 để tính toán và đề xuất công thức (3) [4]. Công thức (3) cho trị số của Y lớn hơn so với công thức (1). Các tác giả khác như Peterka (1984) [3], Bretz (1988), Hager (1992) cũng đã công bố các nghiên cứu của mình dựa trên kết quả thí nghiệm trong máng kính [1,5]. Peterka đề nghị sử dụng công thức (1) để tính toán độ sâu sau nước nhảy và cho rằng độ sâu hạ lưu ít ảnh hưởng đến nước nhảy khi số Fr_1 dao động trong khoảng từ 4,5 đến 9 [3]. Công thức Hager (1992), số (4), có dạng tương tự như (1), được áp dụng khi số $Fr_1 > 2$.

Sarma – Newnham (1973):
$$Y = \frac{h_2}{h_1} = \frac{1}{2}\left(\sqrt{1+10,4Fr_1^2}-1\right) \quad (3)$$

Hager (1992):
$$Y = \sqrt{2}Fr_1 - 0,5 \quad (4)$$

Trên cơ sở nghiên cứu thực nghiệm, Harry Edmar Schulz và các cộng sự (2015) đã đề xuất công thức (5) áp dụng trong trường hợp $2,26 < Fr_1 < 15,96$ [6]. Tuy nhiên, các công thức thực nghiệm cho kết quả khác nhau và đều có sai số tương đối khá cao (xem Bảng 4).

Schulz (2015):
$$Y = 1,29Fr_1 - 0,116 \quad (5)$$



Hình 1. Các độ sâu trước và sau nước nhảy.

Nghiên cứu của Hager và các cộng sự [7] cho thấy, khi số Fr_1 lớn hơn 5,5 và độ sâu h_1 nhỏ, độ sâu sau nước nhảy h_2 sẽ có trị số nhỏ hơn so với tính toán theo công thức (1). Công thức này có sai số tương đối từ 0,1% đến 12,2%. Công thức (5) có sai số lớn nhất là 5,4% [6].

Đó là do tính nhớt của chất lỏng và độ nhám bề mặt lòng dẫn đã gây ra lực ma sát cản trở dòng chảy, làm giảm độ sâu sau nước nhảy so với khi bỏ qua ma sát. Vì vậy, cần xây dựng các mô hình toán để xác định độ sâu sau nước nhảy một cách chính xác, đơn giản và phù hợp với thực tế.

Cho đến nay, các thuật toán học máy (ML) đã và đang được sử dụng rộng rãi trong việc giải quyết nhiều vấn đề thực tế phức tạp, bao gồm tài nguyên nước và xây dựng dân dụng [8,9]. Các mô hình ML thuộc nhóm các mô hình dựa trên mối quan hệ giữa các dữ liệu đầu vào và đầu ra để đưa ra dự báo. Các thuật toán này phục vụ cho bài toán hồi quy và học máy có giám sát. Các mô hình này được ứng dụng để tính toán các đặc trưng của nước nhảy từ năm 2012 [10], kết hợp với mô hình vật lý để giải quyết các bài toán thực tế phức tạp [11,12]. Các mô hình ML như ANFIS (adaptive neuro-fuzzy inference system), ANFIS-PSO (ANFIS-particle swarm optimization), LASSO (least absolute shrinkage and selection operator) đã được áp dụng để tính toán tỷ số hai độ sâu liên hiệp của nước nhảy [12]. Ngoài ra, các mô hình ML như Rừng cây ngẫu nhiên (Random Forest - RT), Tăng cường độ dốc (Gradient Boosting - GB), Cây bổ sung (Extra Trees - ET) đã được sử dụng nhiều để dự báo mực nước mặt và nước ngầm [13], Máy Vector hỗ trợ (Support Vector Machine – SVM) và Tăng cường thích ứng (Adaptive Boosting – Ada) được sử dụng để giải quyết các bài toán thủy lực [14,15]. Các nghiên cứu công bố ở Việt Nam cho thấy rằng, mô hình RT và SVM đã được sử dụng để dự báo chất lượng nước dưới đất ở thành phố Hội An. Tuy nhiên, độ chính xác của kết quả dự báo không cao [16]. Các mô hình ML như: RT, ET, SVM, GB đã được sử dụng để tính toán chiều dài nước nhảy, cho kết quả rất tốt [17,18].

Từ những phân tích trên đây, nghiên cứu này sẽ thiết lập và kiểm định sáu mô hình dựa trên thuật toán ML để xác định tỷ số độ sâu liên hiệp của nước nhảy, áp dụng cho kênh đáy bằng có mặt cắt ngang hình chữ nhật. Các mô hình: RT, Ada, GB, ET, SVM và Tăng cường tốc độ (Cat Boosting – CB) sẽ được xây dựng. Kết quả tính toán của sáu mô hình này sẽ được so sánh với nhau và với bốn công thức kinh nghiệm nhằm đánh giá hiệu quả của từng mô hình.

2. CÁC DỮ LIỆU CẦN THIẾT VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Các dữ liệu cho mô hình ML

Các dữ liệu đầu vào và đầu ra nhằm mục đích huấn luyện và kiểm định các mô hình ML được thu thập từ các công bố của Hager và các cộng sự [7,19]. Trong các nghiên cứu của Hager và các cộng sự, bốn chuỗi thí nghiệm trong máng chữ nhật nằm ngang có chiều rộng 500 mm và 98 mm đã được thực hiện. Đáy máng và tường bên trái làm bằng nhựa PVC, tường bên phải làm bằng kính để dễ dàng quan sát nước nhảy. Do đó, độ nhám tuyệt đối của lòng dẫn mô hình là e = 0,005 mm [7]. Các thí nghiệm được tiến hành ở trong phòng, nhiệt độ từ 16 đến 18 ^{o}C, do đó hệ số nhớt động học của nước là $\upsilon = 1{,}1*10^{-6}$ m^2/s. Các mẫu thí nghiệm đảm bảo mô phỏng đầy đủ các đặc trưng thủy động lực học của nước nhảy ngoài thực tế như: số Froude, số Reynolds, vận tốc, trước nước nhảy, phù hợp với điều kiện Việt Nam. Sau khi tiến hành xử lý dữ liệu để loại bỏ nhiễu, tổng cộng 144 bộ số liệu thí nghiệm của Hager và các cộng sự đã được sử dụng cho nghiên cứu này. Các bộ số liệu này được chia làm hai phần để phục vụ các mô hình ML, phần thứ nhất gồm 115 mẫu (khoảng 80% bộ dữ liệu) dùng cho việc huấn luyện mô hình (training), phần thứ hai gồm 29 mẫu (hơn 20% bộ dữ liệu) nhằm mục đích kiểm định mô hình (testing). Thuật toán ML sẽ chọn ngẫu nhiên 29 bộ số liệu, được dùng chung cho việc kiểm định tất cả các mô hình nhằm đảm bảo tính khách quan. Việc lựa chọn này không phụ thuộc vào ý muốn của người sử dụng mô hình.

Thông số thủy lực của các chuỗi thí nghiệm được thống kê trong Bảng 1. Trong đó, số Reynolds Re_1 tại mặt cắt trước nước nhảy được tính theo các công thức (6), số Reynold hiệu chỉnh tại mặt cắt trước nước nhảy Re_1^* tính theo công thức (7).

$$\mathrm{Re}_1 = \frac{4V_1R_{h1}}{\upsilon} \tag{6}$$

$$\mathrm{Re}_1^* = \frac{V_1h_1}{\upsilon} \tag{7}$$

Trong đó: R_{h1} – bán kính thủy lực tại mặt cắt trước nước nhảy; υ - hệ số nhớt động học của nước. Trong Bảng 1: e - độ nhám tuyệt đối của lòng dẫn; b – chiều rộng máng thí nghiệm.

Bảng 1. Các thông số cơ bản của các thí nghiệm.

Máng thí nghiệm	$\omega = h_1/b$	Fr_1	Re_1 x 10^{-5}	Re_1^* x 10^{-5}	e/h1
1	0,012	6,56-15,31	0,35-0,71	0,09-0,18	0,0008
	0,019	5,24-11,83	0,57-1,11	0,15-0,29	0,0005
	0,03	4,70-8,60	0,95-1,69	0,25-0,45	0,0003
b_1=500mm	0,038	3,26-7,87	0,87-2,14	0,25-0,57	0,0003
	0,057	3,33-5,88	1,61-2,90	0,45-0,81	0,0002
	0,078	3,34-4,36	2,50-3,29	0,72-0,95	0,0001
2	0,1	3,39-10,87	0,30-1,10	0,09-0,33	0,0005
	0,12	4,80-8,98	0,58-1,17	0,18-0,37	0,0004
	0,19	4,01-8,26	0,81-1,81	0,28-0,63	0,0003
	0,26	4,73-6,92	1,41-2,11	0,53-0,80	0,0002
b_2=98mm	0,36	3,49-5,83	1,49-2,53	0,64-1,09	0,0001
	0,39	3,60-5,55	1,69-2,67	0,74-1,18	0,0001
	0,55	2,69-4,56	1,78-3,13	0,92-1,64	0,0001
	0,65	2,45-4,15	1,94-3,43	1,10-2,00	0,0001
3	0,055-0,068	3,96-11,12	1,87-7,04	0,52-1,99	0,0002-0,0001
b_3=500mm	0,098-0,109	2,26-8,56	2,32-10,27	0,69-3,12	0,0001
4	0,01	4,33-15,96	0,19-0,54	0,05-0,14	0,001
	0,021	2,88-11,37	0,34-1,21	0,09-0,31	0,0005
	0,024	2,94-9,91	0,44-1,34	0,12-0,35	0,0004
b_4=500mm	0,048	3,06-6,71	1,15-2,53	0,32-0,69	0,0002
	0,072	2,38-5,38	1,75-3,63	0,50-1,04	0,0001

2.2. Áp dụng phương trình động lượng và Định lý Pi của Buckingham

Các phương trình động lượng và liên tục đã được áp dụng cho đoạn dòng chảy 1-2 trong phạm vi nước nhảy (xem Hình 1) nhằm xác định mối liên hệ giữa tỷ số các độ sâu nước nhảy với các yếu tố thủy lực khác. Kết quả tính toán cho thấy rằng, độ sâu sau nước nhảy h_2 phụ thuộc vào độ sâu h_1 và vận tốc trung bình V_1 tại mặt cắt trước nước nhảy; chiều rộng và độ nhám lòng dẫn; khối lượng riêng ρ và hệ số nhớt μ của chất lỏng; gia tốc trọng trường. Mối liên hệ vật lý này có thể thấy qua phương trình (8).

$$h_2 = f(h_1, V_1, b, e, \rho, \mu, g) \tag{8}$$

Trong phương trình (8) có 8 biến số và đơn vị đo của chúng chứa đủ 3 thứ nguyên cơ bản M, L, T. Do đó, theo Định lý π-Buckingham sẽ tồn tại 5 hàm π thay thế cho 8 biến số trên và các hàm π có mối liên hệ với nhau như trong phương trình (9).

$$\pi_1 = \Phi(\pi_2, \pi_3, \pi_4, \pi_5) \quad (9)$$

Để tìm 5 hàm π này, 3 biến lặp lại trong các hàm π được chọn là h_1, V_1, ρ; 5 biến không lặp lại sẽ là h_2, e, b, μ, g. Kết quả giải một hệ 5 phương trình thu được 5 hàm Π như sau:

$$\boldsymbol{\pi_1 = h_2/h_1 ; \pi_2 = Fr_1 ; \pi_3 = Re_1^* ; \pi_4 = e/h_1 ; \pi_5 = h_1/b}$$

Như vậy, mối liên hệ giữa tỷ số hai độ sâu nước nhảy với các yếu tố thủy lực khác được biểu thị bằng phương trình (10). Phương trình này được sử dụng để xác định các đầu vào và đầu ra của mô hình ML. Có bốn đầu vào là các hàm π từ 2 đến 5 và một đầu ra là hàm π_1.

$$Y = \frac{h_2}{h_1} = \Phi\left(Fr_1, \mathrm{Re}_1^*, \frac{e}{h_1}, \frac{h_1}{b}\right) \quad (10)$$

2.3. Các thuật toán ML

2.3.1. Mô hình Rừng cây ngẫu nhiên (RT)

Mô hình RT có độ chính xác cao, ít bị hiện tượng quá khớp (overfitting) nên được sử dụng rộng rãi trong bài toán dự báo của học máy có giám sát. Mô hình này tạo ra nhiều cây quyết định (Decision Tree - DT) mà mỗi DT được huấn luyện dựa trên nhiều mẫu con khác nhau và kết quả dự báo là giá trị trung bình thu được từ toàn bộ các DT [20]; các DT là hoàn toàn độc lập với nhau. Do đó, kết quả dự báo sẽ không bị sai lệch do các DT đều sử dụng chung một bộ dữ liệu huấn luyện. Mô hình dự báo cho kết quả là giá trị trung bình của các dự báo từ những mô hình con theo phương trình (11).

$$\hat{y}_j = \frac{1}{K}\sum_{i=1}^{K} \hat{y}_j^{(i)} \quad (11)$$

Trong đó: $\hat{y}_j^{(i)}$ là dự báo của quan sát thứ j từ mô hình thứ i, $\hat{y}_j^{(i)} = f_i(x_j)$; x_j là giá trị véc tơ đầu vào; f_i là hàm dự báo của mô hình thứ i; K là số lượng các DT.

2.3.2. Mô hình Tăng cường thích ứng (Ada)

Ada được sử dụng chủ yếu để tăng hiệu suất của các mô hình phân loại yếu. Nguyên tắc chính của mô hình Ada là xem xét từng mẫu trong tập dữ liệu đầu vào dựa trên các lỗi của lần lặp trước đó. Mô hình này sử dụng liên tiếp các mô hình phân loại yếu để điều chỉnh lại trọng số cho các quan sát. Việc điều chỉnh trọng số sau mỗi lần lặp sẽ giúp bộ học yếu tiếp theo tập trung nhiều hơn vào các mẫu bị phân loại sai trước đó. Việc này sẽ tiếp tục diễn ra cho đến khi sai số hội tụ về một giá trị nhỏ nhất [21]. Do tính chất thích ứng của mình, mô hình Ada đạt hiệu quả cao khi dự báo có ranh giới phức tạp giữa các lớp hoặc các bài toán hồi quy phi tuyến. Phương trình hồi quy của Ada được biểu diễn bằng phương trình (12).

$$\hat{y}(x) = \sum_{i=1}^{K} \alpha_i \cdot f_i(x) \quad (12)$$

Trong đó: α_i biểu thị trọng số của cây thứ i, được tính dựa trên sai số của cây đó; x là giá trị véc tơ đầu vào; f_i là hàm dự báo của cây thứ i; K là số lượng các cây.

2.3.3. Mô hình Tăng cường độ dốc (GB)

GB là một thuật toán hiện đại được phát triển dựa trên Ada. GB tương tự như Ada, huấn luyện liên tiếp các mô hình yếu [22]. Thuật toán GB kết hợp các DT, nhưng các DT không hoàn toàn độc lập mà có sự phụ thuộc theo chuỗi. Mô hình GB không sử dụng mẫu tái lặp để

tạo dữ liệu huấn luyện mà sử dụng ngay dữ liệu gốc để huấn luyện. Mô hình này không cố gắng khớp giá trị của biến mục tiêu mà tìm cách khớp giá trị sai số của mô hình trước đó. Bằng cách khớp trên những DT có kích thước rất nhỏ, hàm dự báo sẽ dần được cải thiện trong vùng không dự báo tốt. Kết quả dự báo từ chuỗi mô hình là tổng hợp kết quả của các mô hình con, được trình bày trong phương trình (13).

$$\hat{y}(x) = \sum_{b=1}^{K} \lambda \hat{f}_b(x) \tag{13}$$

Trong đó: $\hat{y}(x)$ là hàm dự báo từ thuật toán GB; x là ma trận đầu vào; $\hat{f}_b(x)$ là hàm dự báo của mô hình thứ b trong chuỗi mô hình dự báo; λ là hệ số; K là số lượng cây.

2.3.4. Mô hình Tăng cường tốc độ (CB)

CB là một thuật toán của tăng cường độ dốc (GB), có thể xử lý dữ liệu phân loại một cách tự nhiên mà không cần mã hóa rõ ràng dưới dạng one-hot hoặc mã hóa nhãn [23]. Ưu điểm chính của CB là khả năng xử lý các biến phân loại mà không cần tiền xử lý một cách thủ công nhờ việc chuyển đổi các biến thành các giá trị bằng số thông qua các kỹ thuật khác nhau như mã hóa one-hot và mã hóa trung bình. CB cũng cung cấp hỗ trợ tích hợp để xử lý các giá trị bị thiếu. Trong bài toán dự báo, phương trình của CB có thể được mô tả như sau:

$$\hat{y}(x) = \sum_{i=1}^{K} \eta_i \cdot f_i(x) \tag{14}$$

Trong đó: η_i biểu thị tốc độ học được nhân với đóng góp của cây thứ i; x là giá trị véc tơ đầu vào; f_i là hàm dự báo của cây thứ i; K là số lượng các cây.

2.3.5. Mô hình Cây bổ sung (ET)

ET xây dựng một tập hợp các DT chưa được cắt tỉa theo quy trình từ trên xuống một cách cổ điển. Nó phân chia các nút bằng cách chọn các điểm cắt hoàn toàn ngẫu nhiên và sử dụng toàn bộ mẫu huấn luyện để tạo cây. Việc sử dụng mẫu huấn luyện ban đầu đầy đủ, không dùng bản sao mẫu tái lặp giúp làm giảm thiểu sai số [24]. Trong ET, thủ tục tách các thuộc tính số được sử dụng nhiều lần với toàn bộ mẫu huấn luyện ban đầu để tạo ra một mô hình tổng hợp. Các dự báo của các DT được tổng hợp để đưa ra dự báo cuối cùng, bằng cách lấy theo đa số trong các bài toán phân loại hoặc lấy trung bình cộng trong các bài toán dự báo như phương trình (15).

$$\hat{y}_j = \frac{1}{M} \sum_{i=1}^{M} \hat{y}_j^{(i)} \tag{15}$$

Trong đó: $\hat{y}_j^{(i)}$ là kết quả dự báo của quan sát thứ j từ mô hình thứ i, $\hat{y}_j^{(i)} = f_i(x_j)$; x_j là giá trị véc tơ đầu vào; f_i là hàm dự báo của mô hình thứ i; M là số lượng cây.

2.3.6. Mô hình Máy vector hỗ trợ (SVM)

Thuật toán SVM khá hiệu quả trong việc dự báo khi học máy có giám sát. Ưu điểm của SVM là hoạt động tốt với những mẫu dữ liệu có kích thước lớn và thường đạt kết quả vượt trội so với các thuật toán khác của học có giám sát. Khi sử dụng đúng hàm kernel sẽ giúp cải thiện đáng kể kết quả dự báo của SVM. Những hàm kernel phổ biến đã được tích hợp bên trong sklearn gồm có: Kernel RBF dựa trên hàm Gaussian RBF biến đổi phi tuyến; Kernel linear (tuyến tính) là tích vô hướng giữa hai véc tơ; Kernel poly (đa thức) tạo ra một đa thức bậc cao kết hợp giữa hai véc tơ; Kernel Sigmoid dựa trên kernel đa thức, chuyển tiếp qua hàm

tanh và có thể biểu diễn theo hàm sigmoid. Trong bài toán dự báo, thuật toán có tên là SVR (Support Vector Regression), kết quả dự báo được thể hiện bằng phương trình (16).

$$\hat{y}(x) = \omega \cdot \phi(x) + b \tag{16}$$

Trong đó: ω là véc tơ trọng số, b là độ sai lệch; Φ(x) là hàm biến đổi véc tơ đầu vào x thông qua hàm kernel.

2.4. Phương pháp đánh giá sai số

Nghiên cứu này sử dụng các giá trị của các chỉ số thống kê như: Sai số tuyệt đối trung bình (MAE), Sai số căn quân phương (RMSE), Sai số tương đối trung bình (MAPE), Hệ số Nash (NSE) [25], để đánh giá sai số của kết quả dự báo từ các mô hình ML. Những chỉ số thống kê này thường được sử dụng trong các mô hình thủy lực để đánh giá mối tương quan giữa dữ liệu thí nghiệm và kết quả tính toán. NSE được dùng để đo lường mối quan hệ tuyến tính giữa các giá trị thực đo và giá trị dự báo. Các chỉ số trên được tính toán theo các công thức từ (17) đến (20). Mô hình ML cho kết quả có độ chính xác cao khi giá trị của các sai số nhỏ, gần bằng 0 và giá trị của NSE lớn, gần bằng 1.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left(\left|P_i - O_i\right|\right) \tag{17}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(P_i - O_i\right)^2} \tag{18}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|O_i - P_i\right|}{O_i} \tag{19}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}\left(O_i - P_i\right)^2}{\sum_{i=1}^{n}\left(O_i - \overline{O}_i\right)^2} \tag{20}$$

Trong đó: O_i - trị số thực đo thứ i; $\overline{O}_i$ - trị số thực đo trung bình thứ i; P_i - trị số dự báo tương ứng thứ i; n - số lần tính.

3. LỰA CHỌN CÁC THAM SỐ MÔ HÌNH VÀ ĐÁNH GIÁ ẢNH HƯỞNG CỦA CÁC BIẾN ĐẦU VÀO

3.1. Chọn lựa các siêu tham số của mô hình ML

Các siêu tham số của mô hình ML đóng vai trò quyết định trong việc nâng cao độ chính xác của kết quả dự báo. Tuy nhiên, không có quy định hay hướng dẫn nào cho việc lựa chọn các siêu tham số này. Các giá trị ban đầu của các tham số có thể tìm được nhờ chiến lược tìm kiếm bằng lưới (grid search). Một công cụ được tích hợp trong thư viện phần mềm mã nguồn mở Keras và Scikit-learn cũng như API cấp cao của TensorFlow 2. Sau đó, các tham số tiếp tục được điều chỉnh bằng phương pháp thử dần cho đến khi kết quả dự báo đạt độ chính xác cao nhất. Đây là một phương pháp được sử dụng rộng rãi nhằm cải thiện độ chính xác và độ tin cậy của kết quả dự báo.

Nghiên cứu này đã sử dụng ngôn ngữ lập trình Python 3.7 và một số thư viện như Numpy, Pandas, Matplotlib để xây dựng sáu mô hình ML, quản lý dữ liệu và vẽ đồ thị nhằm minh họa kết quả dự báo bằng hình ảnh. Các mô hình ML đã được chạy thử nghiệm nhiều lần

để thu được các siêu tham số phù hợp và tạo ra các mô hình tốt nhất. Bảng 2 dưới đây thống kê các tham số chính đã được lựa chọn cho từng mô hình.

Bảng 2. Siêu tham số của các mô hình ML.

TT	Mô hình	Tham số chính	Khoảng giá trị	Giá trị chọn
1	**CB**	learning_rate iterations depth	0,01; 0,05; 0,1; 0,5; 1 100; 500; 1000 3; 5; 7	0,01 500 3
2	**RF**	n_estimators max_depth max_samples max_features criterion	100; 500; 1000; 3000 1; 2; …; 7; None 0,1; 0,2; 0,3 ...; 1,0 1; 2; 3; 4 'squared_error'; 'friedman_mse'	1000 None 1,0 4 'squared_error'
3	**ET**	n_estimators max_depth max_samples max_features min_samples_split criterion	100; 300; 500; 1000 1; 2; …; 7; None 0,1; 0,2; 0,3; …; 1,0 1; 2; 3; 4 2; 3; 4; …; 14 'squared_error'; 'friedman_mse'	300 None 0,9 4 2 'squared_error'
4	**Ada**	n_estimators learning_rate loss base_estimators	500; 1000; 3000; 5000 0,1; 0,2; 0,3; …; 2,0 'linear'; 'square'; 'exponential' DecisionTreeRegressor(max_depth)	1000 0,1 'linear' 4
5	**GB**	n_estimators learning_rate max_depth subsample	100; 500; 1000; 3000; 5000 0,001; 0,01; 0,1; 1,0; 1,3 1; 2; 3; …; 10 0,1; 0,2; 0,3; …; 1,0	1000 0,1 3 1,0
6	**SVR**	criterion: Kernel C epsilon	‘linear’; ‘poly’; ‘rbf’; ‘sigmoid’ 0,01; 0,1; 1; 10; 100 0,01; 0,1; 1	'linear' 10 1

3.2. Đánh giá ảnh hưởng của các biến đầu vào mô hình

Trong Bảng 2, các siêu tham số có ảnh hưởng nhiều nhất đến kết quả dự báo của mỗi mô hình được đặt ở dòng đầu tiên. Các tham số này có độ nhạy cao, ảnh hưởng trực tiếp đến hiệu suất của mô hình. Các tham số có ảnh hưởng ít hơn được đặt ở dòng cuối cùng.

Nhằm thấy rõ ảnh hưởng của từng nhóm không thứ nguyên đến kết quả dự báo của các mô hình ML, tác giả đã sử dụng hàm Correl trong phần mềm Excel để tính toán hệ số tương quan. Nó cho biết mối tương quan giữa tỷ số độ sâu nước nhảy Y với lần lượt: h_1/b, e/h_1, Re_1^* và Fr_1. Theo thứ tự lần lượt, hệ số tương quan của Y với số Fr_1 là 0,999; với e/h_1 là 0,732; với h_1/b là -0,39 và với số Re_1^* là -0,19.

Kết quả tính toán cho thấy rằng, số Froude Fr_1 có ảnh hưởng lớn nhất đến kết quả dự báo, vượt trội so với các biến khác, sau đó đến độ nhám tương đối e/h_1 rồi đến độ rộng tương đối h_1/b, cuối cùng là số Re_1^*. Kết quả này sẽ giúp việc sắp xếp thứ tự các đầu vào mô hình hợp lý nhằm tăng hiệu xuất mô hình.

4. KẾT QUẢ NGHIÊN CỨU VÀ THẢO LUẬN

Sáu mô hình ML đã được phát triển và kiểm tra qua hai giai đoạn: huấn luyện và kiểm định. Mục này trình bày kết quả tính toán các chỉ số thống kê trong các giai đoạn kể trên.

4.1. Kết quả huấn luyện mô hình

Các mô hình ML với bộ tham số đã lựa chọn được huấn luyện bằng tập dữ liệu huấn luyện. Hiệu suất của mỗi mô hình trong giai đoạn huấn luyện được đánh giá bằng các chỉ số thống kê MAE, RMSE, NSE và MAPE. Kết quả đánh giá và so sánh giữa các mô hình dựa trên các chỉ số thống kê được trình bày trong Bảng 3 và Hình 2. Mô hình SVR có kết quả huấn luyện tốt nhất, với hệ số Nash đạt 0,999 và sai số tương đối trung bình MAPE là 1,13 %, sau đó đến mô hình ET với hệ số Nash đạt 0,991 và sai số là 2,16 %. Tiếp theo là GB, RF, Ada; CB có kết quả huấn luyện thấp nhất với sai số lên đến 4,75 %. Tất cả các mô hình ML đều đạt hệ số Nash trên 0,95, chứng tỏ kết quả huấn luyện rất tốt và các mô hình đáng tin cậy.

Bảng 3. Kết quả huấn luyện sáu mô hình ML.

TT	Mô hình	MAE	RMSE	NSE	MAPE
1	**RF**	0,311	0,668	0,975	2,85%
2	**Ada**	0,292	0,695	0,962	2,71%
3	**CB**	0,437	0,798	0,962	4,75%
4	**GB**	0,25	0,573	0,977	2,38%
5	**ET**	0,202	0,407	0,991	2,16%
6	**SVR**	0,091	0,118	0,999	1,13%

4.2. Kết quả kiểm định mô hình và thảo luận

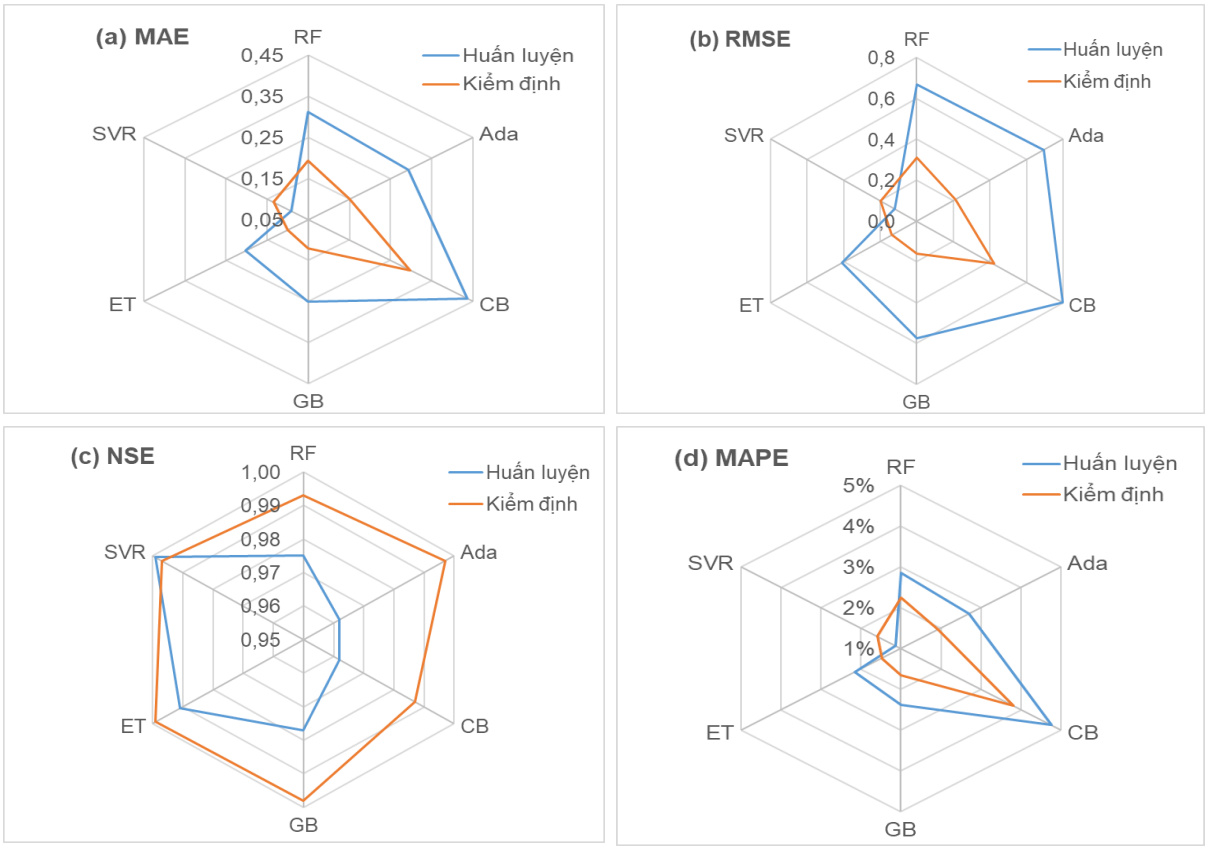
Kết thúc giai đoạn huấn luyện, các mô hình ML bắt đầu được kiểm định bằng bộ dữ liệu kiểm định. Hình 2 so sánh kết quả huấn luyện và kiểm định mô hình, cho thấy trong giai đoạn kiểm định, mô hình CB vẫn có kết quả thấp nhất. Mặc dù có kết quả huấn luyện cao nhất, nhưng mô hình SVR lại có kết quả kiểm định thấp hơn ET và thấp hơn khi huấn luyện. Các chỉ số thống kê khi kiểm định SVR đều thua kém khi huấn luyện. Kết quả này trái ngược với năm mô hình còn lại, khi mà kết quả kiểm định của ET, GB, Ada, RF và CB đều tốt hơn huấn luyện. Việc này chứng tỏ hiệu suất cao của năm mô hình trên trong giai đoạn kiểm định.

Bảng 4 thống kê kết quả kiểm định sáu mô hình ML và bốn công thức kinh nghiệm. Đáng mừng là tất cả các mô hình ML đều có hệ số NSE trên 0,987, chứng tỏ kết quả tính có độ chính xác rất cao. Bốn mô hình ET, GB, SVR, Ada có kết quả kiểm định tốt hơn các công thức kinh nghiệm. Hình 3 thể hiện so sánh số liệu thực đo trong phòng thí nghiệm với kết quả dự báo của sáu mô hình ML. Các điểm trên Hình 3 đều bám sát đường phân giác của góc 90^o, cho biết mô hình có hiệu suất cao và độ chính xác của kết quả dự báo rất ấn tượng.

Bảng 4. Kết quả kiểm định mô hình ML và công thức kinh nghiệm.

TT	Mô hình	MAE	RMSE	NSE	MAPE
1	**RF**	0,192	0,309	0,993	2,24%
2	**Ada**	0,149	0,213	0,997	1,93%
3	**CB**	0,297	0,421	0,987	3,80%
4	**GB**	0,121	0,159	0,998	1,66%
5	**ET**	0,100	0,137	0,999	1,47%
6	**SVR**	0,135	0,197	0,997	1,59%
7	**Belanger**	0,462	0,342	0,970	5,04%
8	**Sarma**	1,735	2,003	0,710	20,75%
9	**Hager**	0,445	0,636	0,971	4,73%
10	**Schulz**	0,148	0,208	0,996	1,95%

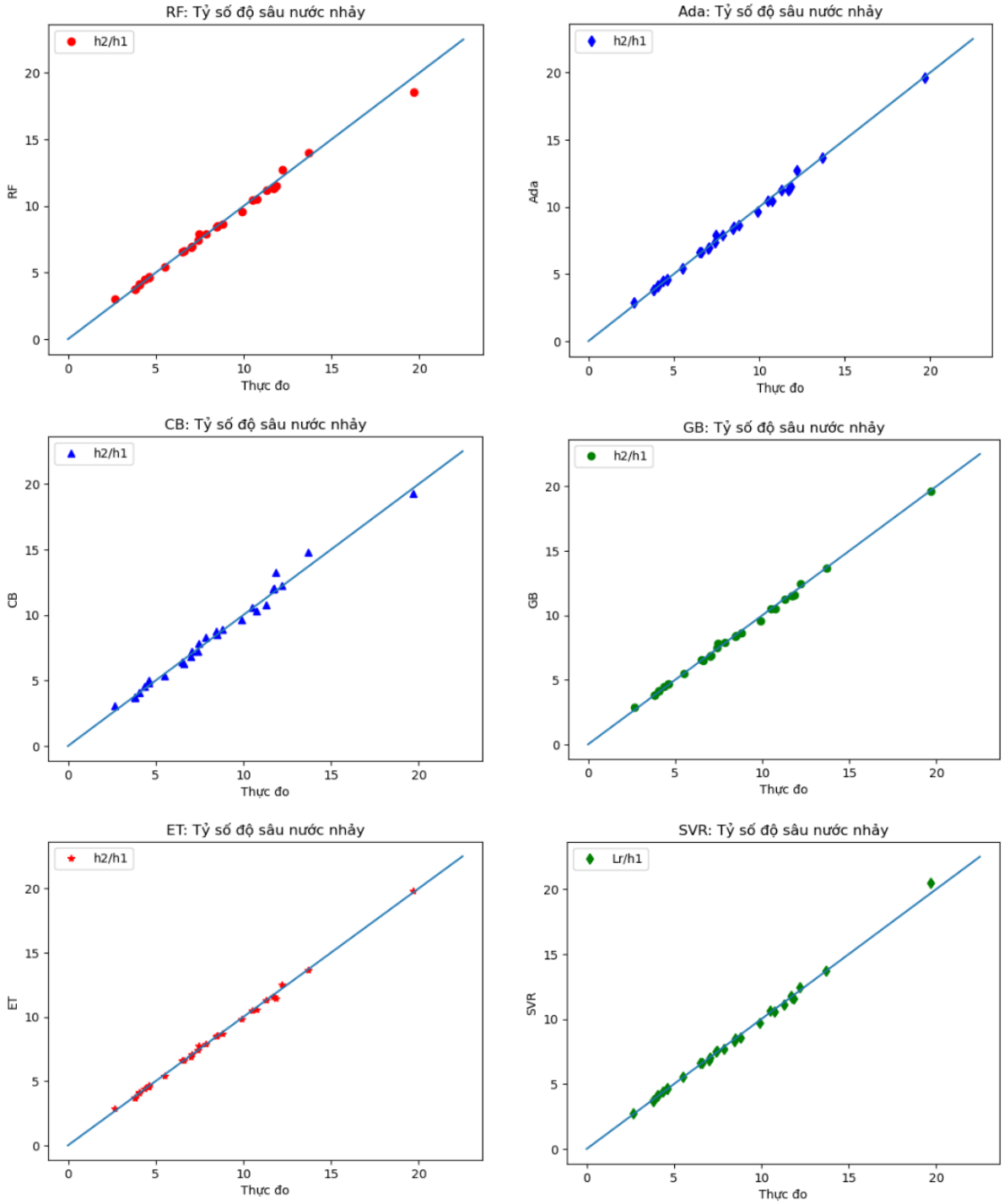
Bảng 4 và Hình 3 cho thấy rằng, mô hình ET có kết quả chính xác nhất, với hệ số NSE đạt 0,999 và sai số tương đối MAPE dưới 1,5 %. Tiếp sau ET là GB, SVR, Ada, RF, CB theo thứ tự độ chính xác giảm dần. Mô hình RF và CB có độ chính xác thấp hơn công thức Schulz (2015) nhưng cao hơn ba công thức còn lại. Trong số các công thức kinh nghiệm, công thức Sarma có sai số lớn nhất với MAPE gần 21%, sau đó đến Belanger và Hager. Công thức Schulz có sai số nhỏ nhất, với trị số của MAPE là 1,95% và hệ số NSE đạt 0,996. Hình 4 chỉ ra rằng, các công thức Belanger, Sarma và Hager cho kết quả tính toán lớn hơn số liệu thực đo, đặc biệt là công thức Sarma có sai số tăng mạnh khi số Fr_1 lớn. Nguyên nhân dẫn đến sai số là các ảnh hưởng của lực ma sát đã không được tính toán đầy đủ.



Hình 2. So sánh kết quả huấn luyện và kiểm định các mô hình ML.

Kết quả kiểm định mô hình và các công thức được thể hiện bằng các đồ thị trong Hình 5. Công thức Schulz (2015) được công bố gần đây nhất cũng là công thức chính xác nhất, vượt trội so với các công thức Belanger, Sarma. Tuy vậy, công thức này vẫn có sai số cao hơn các mô hình ET và GB. Mô hình ET có kết quả tính toán rất ấn tượng với sai số rất nhỏ, kết quả này cũng tương đồng với các nghiên cứu trước đây [15]. Nghiên cứu này cũng cho thấy không phải tất cả các mô hình ML đều tốt hơn các công thức kinh nghiệm, ví dụ RF và CB.

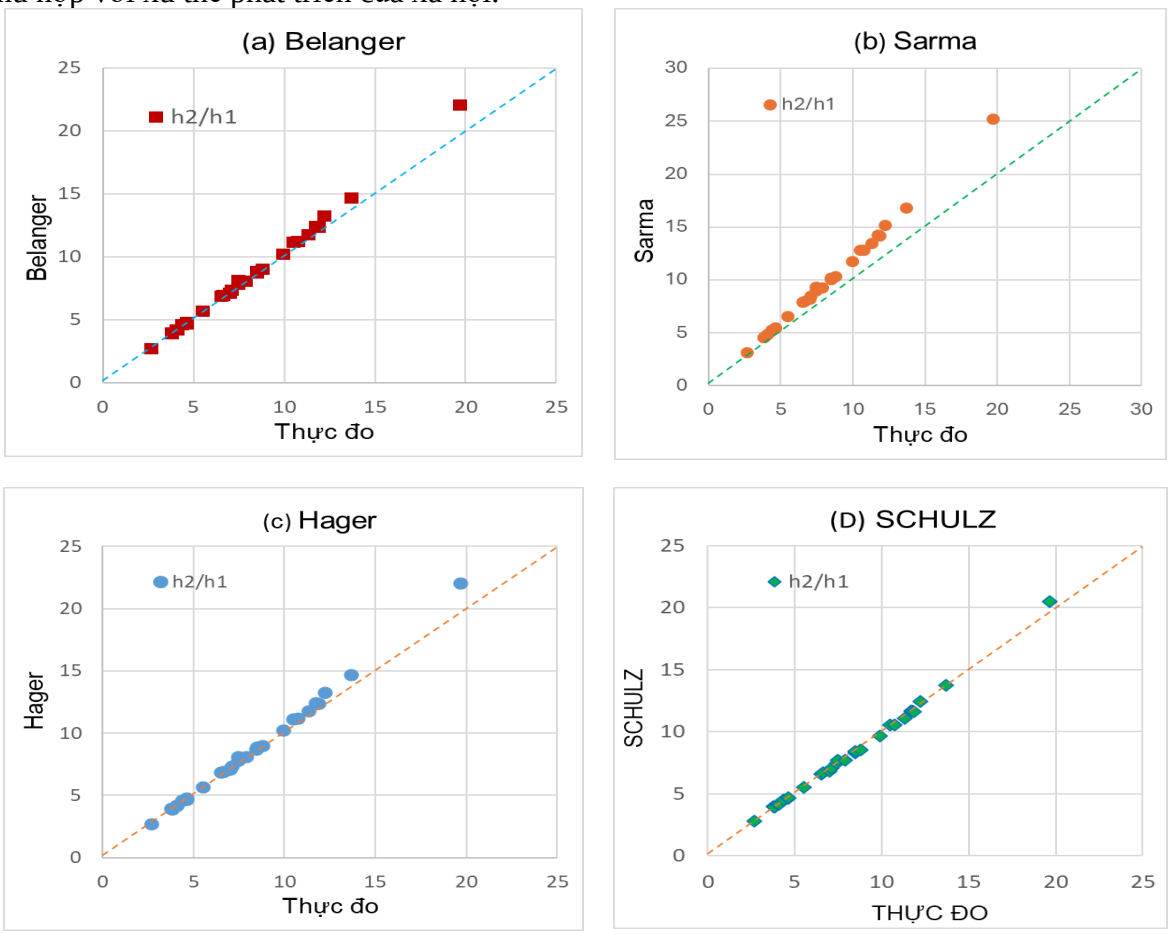
Công thức Schulz có độ chính xác khá cao có thể sánh ngang với một vài mô hình ML như Ada, SVR. Công thức này đã xét đến ảnh hưởng của lực ma sát phát sinh do độ nhám của lòng dẫn và tính nhớt của chất lỏng. Nhưng ảnh hưởng của hai yếu tố trên không thể tách rời nhau và không phân biệt rõ ràng trong công thức Schulz. Các mô hình ML đã phân tách rạch ròi từng yếu tố ảnh hưởng và chú trọng hơn đến những yếu tố ảnh hưởng nhiều nhằm nâng cao hiệu suất dự báo. Ngoài ra, mô hình ML còn xét đến cả ảnh hưởng của chiều rộng lòng dẫn thông qua độ rộng tương đối h_1/b, thông số không có trong công thức Schulz.



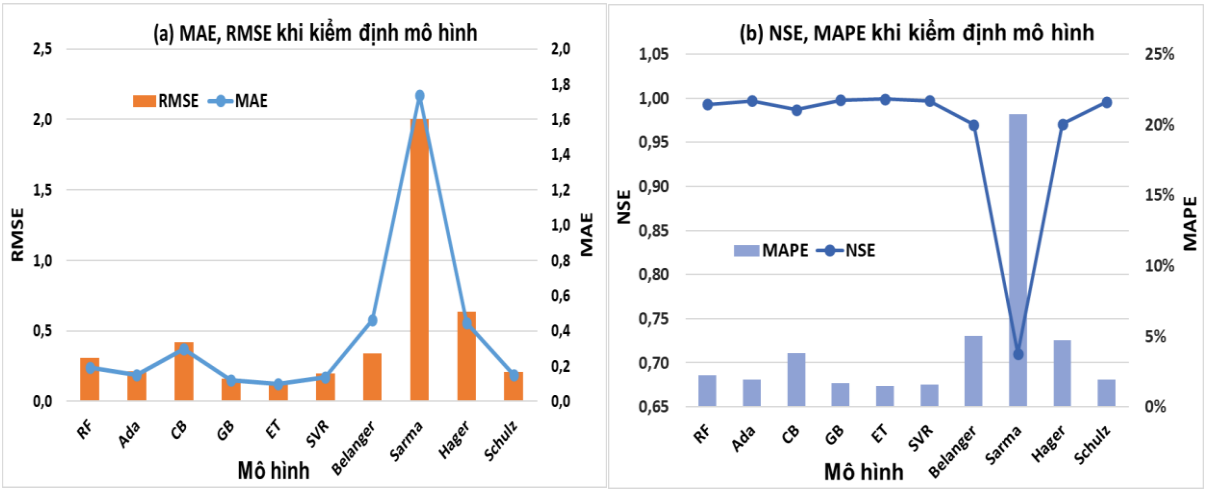
Hình 3. So sánh kết quả kiểm định sáu mô hình ML với số liệu thí nghiệm thực đo.

Nghiên cứu này chứng tỏ các ưu điểm của mô hình ML là: mô hình có kết quả dự báo rất chính xác với hiệu suất cao, có thể đưa nhiều yếu tố thủy lực vào tính toán và chú trọng đến yếu tố ảnh hưởng nhiều, được cập nhật dễ dàng khi có thêm các dữ liệu mới, tốc độ tính toán rất nhanh, dễ tiếp cận. Trong các ưu điểm kể trên, độ chính xác cao và khả năng cập nhật dễ dàng là ưu điểm vượt trội của mô hình ML so với các công thức kinh nghiệm. Hơn nữa, các thuật toán ML ngày càng phát triển và không ngừng được cải tiến để cho ra những mô

hình ML tốt hơn, có độ chính xác cao hơn. Đây là một phương pháp nghiên cứu hiện đại và phù hợp với xu thế phát triển của xã hội.



Hình 4. So sánh kết quả tính của bốn công thức với số liệu thí nghiệm thực đo.



Hình 5. So sánh kết quả kiểm định mô hình với các công thức kinh nghiệm.

Trong nghiên cứu này, các mô hình ML được huấn luyện và kiểm định chỉ bằng các bộ dữ liệu thí nghiệm của Hager và các cộng sự. Bước tiếp theo trong tương lai, khi có các số liệu thí nghiệm mới ở Việt Nam và trên thế giới, các biến đầu vào sẽ được cập nhật và sử dụng để huấn luyện lại mô hình. Việc này sẽ cải thiện độ chính xác của kết quả tính toán và

làm cho mô hình trở nên hiệu quả hơn. Từ đó các mô hình ML có thể được ứng dụng rộng rãi tại Việt Nam.

Tuy nhiên, mô hình ML cũng có những hạn chế như: không kiên định trong các lần dự báo, phụ thuộc vào ý thức của người sử dụng mô hình, cần có dữ liệu đầu vào đủ lớn để đảm bảo độ chính xác cao. Sau mỗi lần chạy, mô hình có thể cho kết quả khác nhau, độ chính xác có thể tăng nhưng cũng có thể giảm. Những hạn chế này có thể được khắc phục bằng phương pháp thử dần, mỗi mô hình cần được chạy nhiều lần để có kết quả tốt nhất, người sử dụng cần thực hành tốt phương pháp tính nhằm đảm bảo kết quả dự báo đúng. Trong sáu mô hình ML, SVR có sự kiên định cao nhất với kết quả dự báo không đổi sau các lần chạy. Để hiểu rõ hơn về vấn đề này cần tiếp tục có các nghiên cứu khác trong thời gian tới.

5. KẾT LUẬN

Nghiên cứu này đã áp dụng một phương pháp mới trong việc xác định độ sâu sau nước nhảy, đó là sử dụng các thuật toán ML, thay cho các công thức kinh nghiệm. Các đầu vào và đầu ra của mô hình toán có được nhờ kết quả tính toán theo Định lý π-Buckingham. Việc này khác với trước đây, một số nghiên cứu đã sử dụng trực tiếp các số liệu thí nghiệm làm đầu vào cho mô hình, bỏ qua các hàm π không thứ nguyên. Ngoài số Froude trước nước nhảy, độ nhám và chiều rộng kênh cũng như độ nhớt của nước đã được đưa vào tính toán, đây là bốn dữ liệu đầu vào mô hình. Ảnh hưởng của các biến đầu vào đến kết quả dự báo cũng được đánh giá kỹ lưỡng nhằm tăng hiệu suất mô hình. Nghiên cứu này đã thiết lập sáu mô hình ML, đánh giá hiệu quả từng mô hình, chỉ ra các ưu nhược điểm của mô hình ML. Sáu mô hình này lần đầu tiên được sử dụng để xác định tỷ số các độ sâu của nước nhảy. Các mô hình ML đã khắc phục được nhược điểm của công thức kinh nghiệm, có thể cập nhật khi được bổ sung dữ liệu mới và đảm bảo độ chính xác cao khi xét đầy đủ các yếu tố ảnh hưởng đến kết quả tính toán. Kết quả tính cho thấy, mô hình ET có độ chính xác cao nhất, sau đó đến GB, SVR, Ada, RF, CB theo thứ tự độ chính xác giảm dần. Các mô hình ET, GB, SVR có kết quả tính chính xác hơn các công thức kinh nghiệm. Vì vậy, có thể sử dụng mô hình ET để tính độ sâu sau nước nhảy trong kênh chữ nhật nằm ngang khi số Froude Fr_1 thay đổi từ 2,26 đến 15,96.

Đóng góp mới của nghiên cứu này là: ứng dụng các mô hình học máy hiện đại để giải quyết bài toán thủy lực; các biến đầu vào và đầu ra của mô hình được chuẩn hóa bằng các nhóm không thứ nguyên. Ngoài ra, nghiên cứu này còn có ý nghĩa thực tế trong việc ứng dụng mô hình toán. Mô hình được sử dụng rất đơn giản, người sử dụng chỉ cần nhập bốn số liệu đầu vào theo đúng định dạng sẽ có kết quả tính toán ngay lập tức.

TÀI LIỆU THAM KHẢO

[1]. WH. Hager, Energy Dissipators and Hydraulic Jump, Kluwer Academic Publishers, Water Science and Technology Library, 8 (1992). https://link.springer.com/book/10.1007/978-94-015-8048-9

[2]. WH. Hager, Hydraulic jump, Energy dissipators, (1995) 60. https://doi.org/10.1201/9780203757512-4

[3]. AJ. Peterka, Hydraulic Design of Stilling Basins and Energy Dissipators, Monogr E, Ed A water Resour Tech Publ USBR, (1984) 240. https://ntrl.ntis.gov/NTRL/dashboard/searchResults/titleDetail/PB95139457.xhtml

[4]. K. V. N. Sarma ; D. A. Newnham, Surface profiles of hydraulic jump for Froude numbers less than four, Water Power, 25 (1973) 139-142.

[5]. N. V. Bretz, Hydraulic Jump Forced by Sill, Ecole Polytechnique Fédérale de Lausanne, 1988.

[6]. HE. Schulz , JD. Nóbrega, ALA. Simões, H. Schulz, Details of Hydraulic Jumps for Design

Criteria of Hydraulic Structures, Hydrodyn - Concepts Exp, (2015). https://doi.org/10.5772/58963

[7]. WH. Hager, R. Bremen, Sequent depths: Le ressaut hydraulique classique: étude des hauteurs conjuguées, J Hydraul Res, 27 (1989) 565-585. https://doi.org/10.1080/00221688909499111

[8]. H. V. Ho, D. H. Nguyen, X. H. Le , G. Lee, Multi-step-ahead water level forecasting for operating sluice gates in Hai Duong, Vietnam, Environ Monit Assess, 194 (2022)1-27. https://doi.org/10.1007/s10661-022-10115-7

[9]. P. N. Huynh, Development of an artificial neural network based-prediction model for bond strength of FRB bars in concrete, Transport Communications Science Journal, 75 (2024)1502-1517. https://doi.org/10.47869/tcsj.75.4.3

[10].M. Naseri , F. Othman, Determination of the length of hydraulic jumps using artificial neural networks, Adv Eng Softw, 48 (2012) 27-31. https://doi.org/10.1016/j.advengsoft.2012.01.003

[11].SL. Brunton , BR. Noack, P. Koumoutsakos, Machine Learning for Fluid Mechanics, Annu Rev Fluid Mech, 52 (2020) 477-508. https://doi.org/10.1146/annurev-fluid-010719-060214

[12].S. Baharvand , A. Jozaghi, R. Fatahi-Alkouhi, S. Karimzadeh , R. Nasiri , B. Lashkar-Ara, Comparative Study on the Machine Learning and Regression-Based Approaches to Predict the Hydraulic Jump Sequent Depth Ratio, Iran J Sci Technol - Trans Civ Eng, 45 (2021) 2719-2732.

[13].K. Kenda , J. Peternelj , N. Mellios , D. Kofinas, M. Čerin, J. Rožanec, Usage of statistical modeling techniques in surface and groundwater level prediction, J Water Supply Res Technol - AQUA, 69 (2020) 248-265. https://doi.org/10.2166/aqua.2020.143

[14]. O. Kisi, P. Khosravinia, MR. Nikpour , H. Sanikhani, Hydrodynamics of river-channel confluence: toward modeling separation zone using GEP, MARS, M5 Tree and DENFIS techniques, Stoch Environ Res Risk Assess, 33 (2019)1089-1107.

[15]. H.V. Ho, Determination of the surface roller length of hydraulic jumps in horizontal rectangular channels using the machine learning method. Stoch Environ Res Risk Assess, 38 (2024) 2539-2562. https://doi.org/10.1007/s00477-024-02697-0

[16].Lê Phước Cường, Ứng dụng mô hình học máy dự báo chất lượng nước dưới đất: điển hình tại khu vực thành phố Hội An, tỉnh Quảng Nam, Tạp chí khoa học và công nghệ - Đại học Đà Nẵng, 20 (2022) 106-110.

[17].Lê Văn Nghị, Ứng dụng thuật toán SVM dự báo chiều dài nước nhảy trên kênh hình thang cân. Tạp chí khoa học và công nghệ thủy lợi, (2023) 43-49.

[18].H.V. Hùng, Ứng dụng phương pháp học máy tính toán chiều dài nước nhảy trong kênh lăng trụ mặt cắt hình chữ nhật, Tạp chí khoa học và công nghệ thủy lợi, 84 (2024)76-86.

[19].W.H. Hager, R. Bremen , N. Kawagoshi, Length of rollerRessaut hydraulique classique: Longueur du rouleau, J Hydraul Res, 28 (1990) 591-608.

[20]. L. Breiman, Random forests, Mach Learn, 45 (2001) 5-32. https://doi.org/10.1023/A:1010933404324

[21].Y. Freund, RE. Schapire, A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, J Comput Syst Sci, 139 (1997) 119-139. https://doi.org/10.1006/jcss.1997.1504

[22].JH. Friedman, Greedy Funtion Approximation: A Gradien Boosting Machine, (1999). https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[23].L. Prokhorenkova , G. Gusev, A. Vorobev, AV. Dorogush, A. Gulin, Catboost: Unbiased boosting with categorical features, Adv Neural Inf Process Syst, 2 (2018) 6638-6648.

[24]. P. Geurts, D. Ernst , L. Wehenkel, Extremely randomized trees, Mach Learn, 63 (2006) 3-42.

[25].JE. Nash, J.V. Sutcliffe, River flow forecasting through conceptual models Part I - A discussion of principles, J Hydrol., 10 (1970) 282-290. https://doi.org/10.1016/0022-1694(70)90255-6